Customer No. 20350

TOWNSEND and TOWNSEND and CREW LLP Two Embarcadero Center, 8th Floor

San Francisco, California 94111-3834 (415) 576-0200

ASSISTANT COMMISSIONER FOR PATENTS BOX PATENT APPLICATION

I hereby certify that this is being deposited with the United States Washington, D.C. 20231 Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above, addressed to: Sir: Transmitted herewith for filing under 37 CFR 1.53(b) is the **Assistant Commissioner for Patents** [X] patent application of Washington, D.C. 20231] continuation patent application of divisional patent application of [] continuation-in-part patent application of By: _ Inventor(s)/Applicant Identifier: Richard STEPHENS, Wayne MITCHELL, Sue KALMAN For: CHLAMYDIA PNEUMONIAE GENOME SEQUENCE This application claims priority from each of the following Application Nos./filing dates: [X] 60/128,606, filed April 8, 1999 and 60/108,279, filed November 12, 1999 the disclosure(s) of which is (are) incorporated by reference. Please amend this application by adding the following before the first sentence: "This application is a [] continuation [] [] continuation-in-part of and claims the benefit of U.S. Application No. 60/_____, filed _____, the disclosure of which is incorporated by reference." Enclosed are: 123 page(s) of specification [**X**] __page(s) of claims [X]1 page of Abstract sheet(s) of [] formal [] informal drawing(s). An assignment of the invention to A [] signed [] unsigned Declaration & Power of Attorney A [] signed [] unsigned Declaration. A Power of Attorney. A verified statement to establish small entity status under 37 CFR 1.9 and 37 CFR 1.27 [] is enclosed [] was filed in the prior application and small entity status is still proper and desired. application. A certified copy of a Information Disclosure Statement under 37 CFR 1.97.

Attorney Docket No. 018941-000411US

"Express Mail" Label No. EL378169002US

November 11, 1999

Date of Deposit:

In view of the Unsigned Declaration as filed with this application and pursuant to 37 CFR §1.53(f), Applicant requests deferral of the filing fee until submission of the Missing Parts of Application.

Notification of change of [] power of attorney [] correspondence address filed in prior application.

DO NOT CHARGE THE FILING FEE AT THIS TIME.

Telephone:

(415) 576-0200

Sequence listing of 203 pages.

Facsimile:

A petition to extend time to respond in the parent application.

(415) 576-0300

Kevin L. Bastian Reg No.: 34,774

Attorneys for Applicant

SF 1043323 v1

[X]

CHLAMYDIA PNEUMONIAE GENOME SEQUENCE

CROSS-REFERENCES TO RELATED APPLICATIONS

The present application is related to 60/128,606, filed April 8, 1999 and 60/108,279, filed November 12, 1998, which are incorporated herein by reference.

STATEMENT AS TO RIGHTS TO INVENTIONS MADE UNDER FEDERALLY SPONSORED RESEARCH AND DEVELOPMENT

10

FIELD OF THE INVENTION

This invention relates to nucleic acids and polypeptides from *Chlamydia* pneumoniae and to their use in the diagnosis, prevention and treatment of diseases associated with *C. pneumoniae*.

15

20

25

BACKGROUND OF THE INVENTION

Chlamydiaceae is a family of obligate intracellular parasite with a tropism for epithelial cells lining the mucus membranes. The bacteria have two morphologically distinct forms, "elementary body" and "reticulate body". The elementary body is the infectious form, and has a rigid cell wall, primarily of cross-linked outer membrane proteins. The reticulate body is the intracellular, metabolically active form. A unique developmental cycle between these two forms characterizes Chlamydia growth.

C. pneumoniae is a human respiratory pathogen that causes acute respiratory disease, and approximately 10% of community-acquired pneumonia. Antibody prevalence studies have shown that virtually everyone is infected with C. pneumoniae at some time, and that reinfection is common. In addition to respiratory disease, studies have shown an association of this organism with coronary artery disease. It has been demonstrated in atherosclerotic lesions of the aorta and coronary arteries by immunocytochemistry and by polymerase chain reaction (Kuo et al. (1993) J Infect Dis 167(4):841-849).

30

Recent reports have further demonstrated the presence of *C. pneumoniae* in the walls of abdominal aortic aneurysms (Juvonen *et al.* (1997) <u>J Vasc Surg</u> **25**(3):499-505). Abdominal aortic aneurysms are frequently associated with atherosclerosis, and inflammation may be an important factor in aneurysmal dilatation.

10

15

20

25

30

C. pneumoniae may play a role in maintaining an inflammation and triggering the development of aortic aneurysms.

Muhlestein et al. (1996) JACC 27:1555-61, reported a differential incidence of *Chlamydia* species within the coronary artery wall of patients with atherosclerosis versus those with other forms of cardiovascular disease. The extremely high rate of possible infection in patients with symptomatic atherosclerotic disease compared to the very low rate in patients with normal coronary arteries or coronary artery disease from chronic transplant rejection provides evidence for a direct link between the atherosclerotic process and *Chlamydia* infection. Because a history of chlamydial infection is so prevalent in the population, the issue of causality remains. On a physiologic and pathologic level, abnormal interactions among endothelial cells, platelets, macrophages and lymphocytes may lead to a cascade of events resulting in acute endothelial damage, thrombosis and repair, chronically leading to the development of atheroma in blood vessels.

C. pneumoniae is related to other Chlamydia species, but the level of sequence similarity is relatively low. Very little is known about the biology of this organism, although it appears to be an important human pathogen. Allelic diversity and structural relationships between specific genes of Chlamydial species is described in Kaltenboeck et al. (1993) J. Bacteriol 175(2):487-502; Gaydos et al. (1992) Infect Immun 60(12):5319-5323; Everett et al. (1997) Int J. Syst Bacteriol 47(2):461-473; and Pudijiatmoko et al. (1997) Int J. Syst Bacteriol 47(2):425-431.

A number of studies have been published describing methods for detection of *C. pneumoniae*, and for distinguishing between Chlamydial species. Such methods include PCR detection (Rasmussen *et al.* (1992) Mol Cell Probes 6(5):389-394; Holland *et al.* (1990) J Infect Dis 162(4):984-987); a simplified polymerase chain reaction-enzyme immunoassay (Wilson et al. (1996) J Appl Bacteriol 80(4):431-438); sequence determination and restriction endonuclease cleavage (Herrmann *et al.* (1996) J Clin Microbiol 34(8):1897-1902).

Antigenic and molecular analyses of different *C. pneumoniae* strains is described in Jantos *et al.* (1997) <u>J Clin Microbiol</u> **35**(3):620-623. Some genes of *C. pneumoniae* have been isolated and sequenced. These include the Gro E operon (Kikuta et al. (1991) <u>Infect Immun</u> **59**(12):4665-4669); the major outer membrane protein Perez *et*

al. (1991) Infect Immun 59(6):2195-2199; the DnaK protein homolog (Kornak et al. (1991) Infect Immun 59(2):721-725); as well as a number of ribosomal and other genes.

5

SUMMARY OF THE INVENTION

This invention provides the genomic sequence of *Chlamydia pneumoniae*. The sequence information is useful for a variety of diagnostic and analytical methods. The genomic sequence may be embodied in a variety of media, including computer readable forms, or as a nucleic acid comprising a selected fragment of the sequence. Such fragments generally consist of an open reading frame, transcriptional or translational control elements, or fragments derived therefrom. Proteins encoded by the open reading frames are useful for diagnostic purposes, as well as for their enzymatic or structural activity.

15

20

25

10

DEFINITIONS

The term "amino acid" refers to naturally occurring and synthetic amino

acids, as well as amino acid analogs and amino acid mimetics that function in a manner similar to the naturally occurring amino acids. Naturally occurring amino acids are those encoded by the genetic code, as well as those amino acids that are later modified, e.g., hydroxyproline, γ-carboxyglutamate, and O-phosphoserine. Amino acid analogs refers to compounds that have the same basic chemical structure as a naturally occurring amino acid, i.e., an α carbon that is bound to a hydrogen, a carboxyl group, an amino group, and an R group., e.g., homoserine, norleucine, methionine sulfoxide, methionine methyl sulfonium Such analogs have modified R groups (e.g., norleucine) or modified peptide backbones, but retain the same basic chemical structure as a naturally occurring amino acid. Amino acid mimetics refers to chemical compounds that have a structure that is different from the general chemical structure of an amino acid, but that functions in a

30

Amino acids may be referred to herein by either their commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single-letter codes.

manner similar to a naturally occurring amino acid.

10

15

20

25

30

"Amplification" primers are oligonucleotides comprising either natural or analogue nucleotides that can serve as the basis for the amplification of a select nucleic acid sequence. They include, e.g., polymerase chain reaction primers and ligase chain reaction oligonucleotides.

"Antibody" refers to an immunoglobulin molecule able to bind to a specific epitope on an antigen. Antibodies can be a polyclonal mixture or monoclonal. Antibodies can be intact immunoglobulins derived from natural sources or from recombinant sources and can be immunoreactive portions of intact immunoglobulins. Antibodies may exist in a variety of forms including, for example, Fv, F_{ab}, and F(ab)₂, as well as in single chains. Single-chain antibodies, in which genes for a heavy chain and a light chain are combined into a single coding sequence, may also be used.

An "antigen" is a molecule that is recognized and bound by an antibody, e.g., peptides, carbohydrates, organic molecules, or more complex molecules such as glycolipids and glycoproteins. The part of the antigen that is the target of antibody binding is an antigenic determinant and a small functional group that corresponds to a single antigenic determinant is called a hapten.

"Biological sample" refers to any sample obtained from a living or dead organism. Examples of biological samples include biological fluids and tissue specimens. Such biological samples can be prepared for analysis of the presence of *C. pneumoniae* nucleic acids, proteins, or antibodies specifically reactive with the proteins.

The term "C. pneumoniae gene" shall be intended to mean the open reading frame encoding specific C. pneumoniae polypeptides, as well as adjacent 5' and 3' non-coding nucleotide sequences involved in the regulation of expression, up to about 2 kb beyond the coding region, but possibly further in either direction. The gene may be introduced into an appropriate vector for extrachromosomal maintenance or for integration into a host genome.

"Conservatively modified variants" applies to both amino acid and nucleic acid sequences. With respect to particular nucleic acid sequences, conservatively modified variants refers to those nucleic acids which encode identical or essentially identical amino acid sequences, or where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. Specifically, degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues

10

15

20

30

(Batzer et al., Nucleic Acid Res. 19:5081 (1991); Ohtsuka et al., J. Biol. Chem. 260:2605-2608 (1985); Rossolini et al., Mol. Cell. Probes 8:91-98 (1994)). Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of conservatively modified variations. Every nucleic acid sequence herein which encodes a polypeptide also describes every possible silent variation of the nucleic acid. One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine, and TGG, which is ordinarily the only codon for tryptophan) can be modified to yield a functionally identical molecule. Accordingly, each silen: variation of a nucleic acid which encodes a polypeptide is implicit in each described sequence.

As to amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a nucleic acid, peptide, polypeptide, or protein sequence which alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservatively modified variant" where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. Such conservatively modified variants are in addition to and do not exclude polymorphic variants, interspecies homologs, and alleles of the invention.

The following groups each contain amino acids that are conservative substitutions for one another:

- 25 1) Alanine (A), Glycine (G);
 - 2) Serine (S), Threonine (T);
 - 3) Aspartic acid (D), Glutamic acid (E);
 - 4) Asparagine (N), Glutamine (Q);
 - 5) Cysteine (C), Methionine (M);
 - 6) Arginine (R), Lysine (K), Histidine (H);
 - 7) Isoleucine (I), Leucine (L), Valine (V); and
 - 8) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

see, e.g., Creighton, Proteins (1984)).

10

15

20

25

30

The terms "identical" or percent "identity," in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same, when compared and aligned for maximum correspondence over a comparison window, as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection. This definition also refers to the complement of a test sequence, which has a designated percent sequence or subsequence complementarity when the test sequence has a designated or substantial identity to a reference sequence. For example, a designated amino acid percent identity of 95% refers to sequences or subsequences that have at least about 95% amino acid identity when aligned for maximum correspondence over a comparison window as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection. Such sequences would then be said to have substantial identity, or to be substantially identical to each other. Preferably, sequences have at least about 70% identity, more preferably 80% identity, more preferably 90-95% identity and above. Preferably, the percent identity exists over a region of the sequence that is at least about 25 amino acids in length, more preferably over a region that is 50-100 amino acids in length.

When percentage of sequence identity is used in reference to proteins or peptides, it is recognized that residue positions that are not identical often differ by conservative amino acid substitutions, where amino acids residues are substituted for other amino acid residues with similar chemical properties (e.g., charge or hydrophobicity) and therefore do not change the functional properties of the molecule. Where sequences differ in conservative substitutions, the percent sequence identity may be adjusted upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well known to those of skill in the art. Typically this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. Thus, for example, where an identical amino acid is given a score of 1 and a non-conservative substitution is given a score of zero, a conservative substitution is given a score between zero and 1. The scoring of conservative substitutions is calculated according to, e.g., the algorithm of Meyers & Miller, *Computer Applic. Biol. Sci.* 4:11-17 (1988) e.g., as implemented in the program PC/GENE (Intelligenetics, Mountain View, California, USA)..

10

15

20

25

30

For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Default program parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence, based on the designated or default program parameters.

A comparison window includes reference to a segment of any one of the number of contiguous positions selected from the group consisting of from 25 to 600, usually about 50 to about 200, more usually about 100 to about 150 in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. Methods of alignment of sequences for comparison are well-known in the art. Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith & Waterman, Adv. Appl. Math. 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, J. Mol. Biol. 48:443 (1970), by the search for similarity method of Pearson & Lipman, Proc. Nat'l. Acad. Sci. USA 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by manual alignment and visual inspection (see, e.g., Ausubel et al., supra).

One example of a useful algorithm is PILEUP. PILEUP creates a multiple sequence alignment from a group of related sequences using progressive, pairwise alignments to show relationship and percent sequence identity. It also plots a tree or dendogram showing the clustering relationships used to create the alignment. PILEUP uses a simplification of the progressive alignment method of Feng & Doolittle, *J. Mol. Evol.* 35:351-360 (1987). The method used is similar to the method described by Higgins & Sharp, *CABIOS* 5:151-153 (1989). The program can align up to 300 sequences, each of a maximum length of 5,000 nucleotides or amino acids. The multiple alignment procedure begins with the pairwise alignment of the two most similar sequences, producing a cluster of two aligned sequences. This cluster is then aligned to the next most related sequence or cluster of aligned sequences. Two clusters of sequences are aligned by a simple extension of the pairwise alignment of two individual sequences. The

10

15

20

25

30

final alignment is achieved by a series of progressive, pairwise alignments. The program is run by designating specific sequences and their amino acid or nucleotide coordinates for regions of sequence comparison and by designating the program parameters. Using PILEUP, a reference sequence is compared to other test sequences to determine the percent sequence identity relationship using the following parameters: default gap weight (3.00), default gap length weight (0.10), and weighted end gaps. PILEUP can be obtained from the GCG sequence analysis software package, e.g, version 7.0 (Devereaux et al., Nuc. Acids Res. 12:387-395 (1984).

Another example of algorithm that is suitable for determining percent sequence identity (i.e., substantial similarity or identity) is the BLAST algorithm, which is described in Altschul et al., J. Mol. Biol. 215:403-410 (1990). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul et al, supra). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues, always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, M=5, N=4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as default parameters a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff & Henikoff, Proc. Natl. Acad. Sci. USA 89:10915 (1989)).

10

15

20

25

30

The BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin & Altschul, Proc. Nat'l. Acad. Sci. USA 90:5873-5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001.

An indication that two nucleic acid sequences or polypeptides are substantially identical is that the polypeptide encoded by the first nucleic acid is immunologically cross reactive with the antibodies raised against the polypeptide encoded by the second nucleic acid, as described below. Thus, a polypeptide is typically substantially identical to a second polypeptide, for example, where the two peptides differ only by conservative substitutions. Another indication that two nucleic acid sequences are substantially identical is that the two molecules or their complements hybridize to each other under stringent conditions, as described below.

Another indication that polynucleotide sequences are substantially identical is if two molecules hybridize to each other under stringent conditions. Stringent conditions are sequence dependent and will be different in different circumstances. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength and pH. The Tm is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Typically stringent conditions for a Southern blot protocol involve hybridizing in a buffer comprising 5x SSC, 1% SDS at 65°C or hybridizing in a buffer containing 5x SSC and 1% SDS at 42°C and washing at 65°C with a 0.2x SSC, 0.1% SDS wash.

A "label" is a composition detectable by spectroscopic, photochemical, biochemical, immunochemical, or chemical means. For example, useful labels include ³²P, fluorescent dyes, electron-dense reagents, enzymes (e.g., as commonly used in an ELISA), biotin, dioxigenin, or haptens and proteins for which antisera or monoclonal antibodies are available.

10

15

20

25

30

The term "nucleic acid" refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form. The term encompasses nucleic acids containing known nucleotide analogs or modified backbone residues or linkages, which are synthetic, naturally occurring, and non-naturally occurring, which have similar binding properties as the reference nucleic acid, and which are metabolized in a manner similar to the reference nucleotides. Examples of such analogs include, without limitation, phosphorothioates, phosphoramidates, methyl phosphonates, chiral-methyl phosphonates, 2-O-methyl ribonucleotides, peptide-nucleic acids (PNAs).

Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses conservatively modified variants thereof (e.g., degenerate codon substitutions) and complementary sequences, as well as the sequence explicitly indicated. The term nucleic acid is used interchangeably with gene, cDNA, mRNA, oligonucleotide, and polynucleotide.

As used herein a "nucleic acid probe or oligonucleotide" is defined as a nucleic acid capable of binding to a target nucleic acid of complementary sequence through one or more types of chemical bonds, usually through complementary base pairing, usually through hydrogen bond formation. As used herein, a probe may include natural (i.e., A, G, C, or T) or modified bases (7-deazaguanosine, inosine, etc.). In addition, the bases in a probe may be joined by a linkage other than a phosphodiester bond, so long as it does not interfere with hybridization. Thus, for example, probes may be peptide nucleic acids in which the constituent bases are joined by peptide bonds rather than phosphodiester linkages. It will be understood by one of skill in the art that probes may bind target sequences lacking complete complementarity with the probe sequence depending upon the stringency of the hybridization conditions. The probes are preferably directly labeled as with isotopes, chromophores, lumiphores, chromogens, or indirectly labeled such as with biotin to which a streptavidin complex may later bind. By assaying for the presence or absence of the probe, one can detect the presence or absence of the select sequence or subsequence.

A labeled nucleic acid probe or oligonucleotide is one that is bound, either covalently, through a linker, or through ionic, van der Waals or hydrogen bonds to a label such that the presence of the probe may be detected by detecting the presence of the label bound to the probe.

10

15

20

25

30

"Pharmaceutically acceptable" means a material that is not biologically or otherwise undesirable, i.e., the material can be administered to an individual along with a *Chlamydia* antigen without causing any undesirable biological effects or interacting in a deleterious manner with any of the other components of the pharmaceutical composition.

The terms "polypeptide," "peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residue is an analog or mimetic of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers.

The phrase "specifically or selectively hybridizing to," refers to hybridization between a probe and a target sequence in which the probe binds substantially only to the target sequence, forming a hybridization complex, when the target is in a heterogeneous mixture of polynucleotides and other compounds. Such hybridization is determinative of the presence of the target sequence. Although the probe may bind other unrelated sequences, at least 90%, preferably 95% or more of the hybridization complexes formed are with the target sequence.

The term "recombinant" when used with reference to a cell, or nucleic acid, or vector, indicates that the cell, or nucleic acid, or vector, has been modified by the introduction of a heterologous nucleic acid or the alteration of a native nucleic acid, or that the cell is derived from a cell so modified. Thus, for example, recombinant cells express genes that are not found within the native (non-recombinant) form of the cell or express native genes that are otherwise abnormally expressed, under expressed or not expressed at all.

The phrase "specifically immunoreactive with", when referring to a protein or peptide, refers to a binding reaction between the protein and an antibody which is determinative of the presence of the protein in the presence of a heterogeneous population of proteins and other compounds. Thus, under designated immunoassay conditions, the specified antibodies bind to a particular protein and do not bind in a significant amount to other proteins present in the sample. Specific binding to an antibody under such conditions may require an antibody that is selected for its specificity for a particular protein. A variety of immunoassay formats may be used to select antibodies specifically immunoreactive with a particular protein and are described in detail below.

10

15

20

25

30

The phrase "substantially pure" or "isolated" when referring to a *Chlamydia* peptide or protein, means a chemical composition which is free of other subcellular components of the *Chlamydia* organism. Typically, a monomeric protein is substantially pure when at least about 85% or more of a sample exhibits a single polypeptide backbone. Minor variants or chemical modifications may typically share the same polypeptide sequence. Depending on the purification procedure, purities of 85%, and preferably over 95% pure are possible. Protein purity or homogeneity may be indicated by a number of means well known in the art, such as polyacrylamide gel electrophoresis of a protein sample, followed by visualizing a single polypeptide band on a polyacrylamide gel upon silver staining. For certain purposes high resolution will be needed and HPLC or a similar means for purification utilized.

DETAILED DESCRIPTION

The present invention provides the nucleotide sequence of the *C. pneumoniae* genome SEQ ID NO: 1 or a representative fragment thereof, in a form which can be readily used, analyzed, and interpreted by a skilled artisan. As used herein, a "representative fragment" of the nucleotide sequence depicted in SEQ ID NO: 1 refers to any portion which is not presently represented within a publicly available database. Preferred representative fragments of the present invention are open reading frames, expression modulating fragments, uptake modulating fragments, and fragments which can be used to diagnose the presence of *C. pneumoniae* in sample. Using the information provided in the present application, together with routine cloning and sequencing methods, one of ordinary skill in the art will be able to clone and sequence all "representative fragments" of interest including open reading frames (ORFs) encoding a large variety of *C. pneumoniae* proteins. A non-limiting identification of such preferred representative fragments is provided in Tables 2 and 3.

Diagnostic use of C. pneumoniae nucleic acids

Hybridization-based assays

Using the nucleic acids disclosed here, one of skill can design nucleic acid hybridization-based assays for the detection of *C. pneumoniae*. Any of a number of well known techniques for the specific detection of target nucleic acids can be used.

Exemplary hybridization-based assays include, but are not limited to, traditional "direct

25

30

5

10

probe" methods such as Southern Blots, dot blots, in situ hybridization (e.g., FISH), PCR, and the like. The methods can be used in a wide variety of formats including, but not limited to substrate- (e.g. membrane or glass) bound methods or array-based approaches as described below. As noted above, this invention also embraces methods for detecting the presence of Chlamydia DNA or RNA in biological samples. These sequences can be used to detect Chlamydia in biological samples from patients suspected of being infected. A variety of methods of specific DNA and RNA measurement using nucleic acid hybridization techniques are known to those of skill in the art (see Sambrook et al., supra).

In situ hy bridization assays are well known (e.g., Angerer (1987) Meth. Enzymol 152: 649). Generally, in situ hybridization comprises the following major steps: (1) fixation of tissue or biological structure to analyzed; (2) prehybridization treatment of the biological structure to increase accessibility of target DNA, and to reduce nonspecific binding; (3) hybridizatic n of the mixture of nucleic acids to the nucleic acid in the biological structure or tissue; (4) post-hybridization washes to remove nucleic acid fragments not bound in the hybridization and (5) detection of the hybridized nucleic acid fragments. The reagent used in each of these steps and the conditions for use vary depending on the particular application.

In a typical in situ hybridization assay, cells are fixed to a solid support, typically a glass slide. If a nucleic acid is to be probed, the cells are typically denatured with heat or alkali. The cells are then contacted with a hybridization solution at a moderate temperature to permit annealing of labeled probes specific to the nucleic acid sequence encoding the protein. The targets (e.g., cells) are then typically washed at a predetermined stringency or at an increasing stringency until an appropriate signal to noise ratio is obtained.

The nucleic acids of this invention are particularly well suited to arraybased hybridization formats. Arrays are a multiplicity of different "probe" or "target" nucleic acids (or other compounds) attached to one or more surfaces (e.g., solid, membrane, or gel). In a preferred embodiment, the multiplicity of nucleic acids (or other moieties) is attached to a single contiguous surface or to a multiplicity of surfaces juxtaposed to each other.

In an array format a large number of different hybridization reactions can be run essentially "in parallel." This provides rapid, essentially simultaneous, evaluation

10

15

20

25

30

of a number of hybridizations in a single "experiment". Methods of performing hybridization reactions in array based formats are well known to those of skill in the art (see, e.g., Pastinen (1997) Genome Res. 7: 606-614; Jackson (1996) Nature Biotechnology 14:1685; Chee (1995) Science 274: 610; WO 96/17958.

Arrays, particularly nucleic acid arrays can be produced according to a wide variety of methods well known to those of skill in the art. For example, in a simple embodiment, "low density" arrays can simply be produced by spotting (e.g. by hand using a pipette) different nucleic acids at different locations on a solid support (e.g. a glass surface, a membrane, etc.).

This simple spotting, approach has been automated to produce high density spotted arrays (*see*, *e.g.*, U.S. Patent No: 5,807,522). This patent describes the use of an automated systems that taps a microcapillary against a surface to deposit a small volume of a biological sample. The process is repeated to generate high density arrays. Arrays can also be produced using oligonucleotide synthesis technology. Thus, for example, U.S. Patent No. 5,143,854 and PCT patent publication Nos. WO 90/15070 and 92/10092 teach the use of light-directed combinatorial synthesis of high density oligonucleotide arrays.

Many methods for immobilizing nucleic acids on a variety of solid surfaces are known in the art. A wide variety of organic and inorganic polymers, as well as other materials, both natural and synthetic, can be employed as the material for the solid surface. Illustrative solid surfaces include, e.g., nitrocellulose, nylon, glass, quartz, diazotized membranes (paper or nylon), silicones, polyformaldehyde, cellulose, and cellulose acetate. In addition, plastics such as polyethylene, polypropylene, polystyrene, and the like can be used. Other materials which may be employed include paper, ceramics, metals, metalloids, semiconductive materials, cermets or the like. In addition, substances that form gels can be used. Such materials include, e.g., proteins (e.g., gelatins), lipopolysaccharides, silicates, agarose and polyacrylamides. Where the solid surface is porous, various pore sizes may be employed depending upon the nature of the system.

In preparing the surface, a plurality of different materials may be employed, particularly as laminates, to obtain various properties. For example, proteins (e.g., bovine serum albumin) or mixtures of macromolecules (e.g., Denhardt's solution) can be employed to avoid non-specific binding, simplify covalent conjugation, enhance

10

15

20

25

30

signal detection or the like. If covalent bonding between a compound and the surface is desired, the surface will usually be polyfunctional or be capable of being polyfunctionalized. Functional groups which may be present on the surface and used for linking can include carboxylic acids, aldehydes, amino groups, cyano groups, ethylenic groups, hydroxyl groups, mercapto groups and the like. The manner of linking a wide variety of compounds to various surfaces is well known and is amply illustrated in the literature.

For example, methods for immobilizing nucleic acids by introduction of various functional groups to the molecules is known (see, e.g., Bischoff (1987) Anal. Biochem., 164: 336-344; Kremsky (1987) Nucl. Acids Res. 15: 2891-2910). Modified nucleotides can be placed on the target using PCR primers containing the modified nucleotide, or by enzymatic end labeling with modified nucleotides. Use of glass or membrane supports (e.g., nitrocellulose, nylon, polypropylene) for the nucleic acid arrays of the invention is advantageous because of well developed technology employing manual and robotic methods of arraying targets at relatively high element densities. Such membranes are generally available and protocols and equipment for hybridization to membranes is well known.

Target elements of various sizes, ranging from 1 mm diameter down to 1 μm can be used. Smaller target elements containing low amounts of concentrated, fixed probe DNA are used for high complexity comparative hybridizations since the total amount of sample available for binding to each target element will be limited. Thus it is advantageous to have small array target elements that contain a small amount of concentrated probe DNA so that the signal that is obtained is highly localized and bright. Such small array target elements are typically used in arrays with densities greater than 10⁴/cm². Relatively simple approaches capable of quantitative fluorescent imaging of 1 cm² areas have been described that permit acquisition of data from a large number of target elements in a single image (see, e.g., Wittrup (1994) Cytometry 16:206-213).

If fluorescently labeled nucleic acid samples are used, arrays on solid surface substrates with much lower fluorescence than membranes, such as glass, quartz, or small beads, can achieve much better sensitivity. Substrates such as glass or fused silica are advantageous in that they provide a very low fluorescence substrate, and a highly efficient hybridization environment. Covalent attachment of the target nucleic acids to glass or synthetic fused silica can be accomplished according to a number of

10

15

20

25

30

known techniques (described above). Nucleic acids can be conveniently coupled to glass using commercially available reagents. For instance, materials for preparation of silanized glass with a number of functional groups are commercially available or can be prepared using standard techniques (see, e.g., Gait (1984) Oligonucleotide Synthesis: A Practical Approach, IRL Press, Wash., D.C.). Quartz cover slips, which have at least 10-fold lower autofluorescence than glass, can also be silanized.

Alternatively, probes can also be immobilized on commercially available coated beads or other surfaces. For instance, biotin end-labeled nucleic acids can be bound to commercially available avidin-coated beads. Streptavidin or anti-digoxigenin antibody can also be attached to silanized glass slides by protein-mediated coupling using e.g., protein A following standard protocols (see, e.g., Smith (1992) Science 258: 1122-1126). Biotin or digoxigenin end-labeled nucleic acids can be prepared according to standard techniques. Hybridization to nucleic acids attached to beads is accomplished by suspending them in the hybridization mix, and then depositing them on the glass substrate for analysis after washing. Alternatively, paramagnetic particles, such as ferric oxide particles, with or without avidin coating, can be used.

A variety of other nucleic acid hybridization formats are known to those skilled in the art. For example, common formats include sandwich assays and competition or displacement assays. Hybridization techniques are generally described in Hames and Higgins (1985) *Nucleic Acid Hybridization, A Practical Approach*, IRL Press; Gall and Pardue (1969) *Proc. Natl. Acad. Sci. USA* 63: 378-383; and John *et al.* (1969) *Nature* 223: 582-587.

Sandwich assays are commercially useful hybridization assays for detecting or isolating nucleic acid sequences. Such assays utilize a "capture" nucleic acid covalently immobilized to a solid support and a labeled "signal" nucleic acid in solution. The sample will provide the target nucleic acid. The "capture" nucleic acid and "signal" nucleic acid probe hybridize with the target nucleic acid to form a "sandwich" hybridization complex. To be most effective, the signal nucleic acid should not hybridize with the capture nucleic acid.

Detection of a hybridization complex may require the binding of a signal generating complex to a duplex of target and probe polynucleotides or nucleic acids. Typically, such binding occurs through ligand and anti-ligand interactions as between a ligand-conjugated probe and an anti-ligand conjugated with a signal.

10

15

20

25

30

The sensitivity of the hybridization assays may be enhanced through use of a nucleic acid amplification system that multiplies the target nucleic acid being detected. Examples of such systems include the polymerase chain reaction (PCR) system and the ligase chain reaction (LCR) system. Other methods recently described in the art are the nucleic acid sequence based amplification (NASBAO, Cangene, Mississauga, Ontario) and O Beta Replicase systems.

Nucleic acid hybridization simply involves providing a denatured probe and target nucleic acid under conditions where the probe and its complementary target can form stable hybrid duplexes through complementary base pairing. The nucleic acids that do not form hybrid duplexes are then washed away leaving the hybridized nucleic acids to be detected, typically through detection of an attached detectable label. It is generally recognized that nucleic acids are denatured by increasing the temperature or decreasing the salt concentration of the buffer containing the nucleic acids, or in the addition of chemical agents, or the raising of the pH. Under low stringency conditions (e.g., low temperature and/or high salt and/or high target concentration) hybrid duplexes (e.g., DNA:DNA, RNA:RNA, or RNA:DNA) will form even where the annealed sequences are not perfectly complementary. Thus specificity of hybridization is reduced at lower stringency. Conversely, at higher stringency (e.g., higher temperature or lower salt) successful hybridization requires fewer mismatches.

One of skill in the art will appreciate that hybridization conditions may be selected to provide any degree of stringency. In a preferred embodiment, hybridization is performed at low stringency to ensure hybridization and then subsequent washes are performed at higher stringency to eliminate mismatched hybrid duplexes. Successive washes may be performed at increasingly higher stringency (e.g., down to as low as 0.25 X SSPE-T at 37°C to 70°C) until a desired level of hybridization specificity is obtained. Stringency can also be increased by addition of agents such as formamide. Hybridization specificity may be evaluated by comparison of hybridization to the test probes with hybridization to the various controls that can be present.

In general, there is a tradeoff between hybridization specificity (stringency) and signal intensity. Thus, in a preferred embodiment, the wash is performed at the highest stringency that produces consistent results and that provides a signal intensity greater than approximately 10% of the background intensity. Thus, in a preferred embodiment, the hybridized array may be washed at successively higher

10

15

20

25

30

stringency solutions and read between each wash. Analysis of the data sets thus produced will reveal a wash stringency above which the hybridization pattern is not appreciably altered and which provides adequate signal for the particular probes of interest.

Methods of optimizing hybridization conditions are well known to those of skill in the art (see, e.g., Tijssen (1993) Laboratory Techniques in Biochemistry and .

Molecular Biology, Vol. 24: Hybridization With Nucleic Acid Probes, Elsevier, N.Y.).

Labeling and detection of nucleic acids.

In a preferred embodiment, the hybridized nucleic acids are detected by detecting one or more labels attached to the sample or probe nucleic acids. The labels may be incorporated by any of a number of means well known to those of skill in the art. Means of attaching labels to nucleic acids include, for example nick translation or endlabeling (e.g. with a labeled RNA) by kinasing of the nucleic acid and subsequent attachment (ligation) of a nucleic acid linker joining the sample nucleic acid to a label (e.g., a fluorophore). A wide variety of linkers for the attachment of labels to nucleic acids are also known. In addition, intercalating dyes and fluorescent nucleotides can also be used.

Detectable labels suitable for use in the present invention include any composition detectable by spectroscopic, photochemical, biochemical, immunochemical, electrical, optical or chemical means. Useful labels in the present invention include biotin for staining with labeled streptavidin conjugate, magnetic beads (*e.g.*, DynabeadsTM), fluorescent dyes (*e.g.*, fluorescein, texas red, rhodamine, green fluorescent protein, and the like, *see. e.g.*, Molecular Probes, Eugene, Oregon, USA), radiolabels (*e.g.*, ³H, ¹²⁵I, ³⁵S, ¹⁴C, or ³²P), enzymes (*e.g.*, horse radish peroxidase, alkaline phosphatase and others commonly used in an ELISA), and colorimetric labels such as colloidal gold (*e.g.*, gold particles in the 40 -80 nm diameter size range scatter green light with high efficiency) or colored glass or plastic (*e.g.*, polystyrene, polypropylene, latex, etc.) beads. Patents teaching the use of such labels include U.S. Patent Nos. 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149; and 4,366,241.

A fluorescent label is preferred because it provides a very strong signal with low background. It is also optically detectable at high resolution and sensitivity through a quick scanning procedure. The nucleic acid samples can all be labeled with a single label, e.g., a single fluorescent label. Alternatively, in another embodiment, different nucleic acid samples can be simultaneously hybridized where each nucleic acid

10

15

20

25

30

sample has a different label. For instance, one target could have a green fluorescent label and a second target could have a red fluorescent label. The scanning step will distinguish cites of binding of the red label from those binding the green fluorescent label. Each nucleic acid sample (target nucleic acid) can be analyzed independently from one another.

Suitable chromogens which can be employed include those molecules and compounds which absorb light in a distinctive range of wavelengths so that a color can be observed or, alternatively, which emit light when irradiated with radiation of a particular wave length or wave length range, *e.g.*, fluorescers.

Desirably, fluorescers should absorb light above about 300 nm, preferably about 350 nm, and more preferably above about 400 nm, usually emitting at wavelengths greater than about 10 nm higher than the wavelength of the light absorbed. It should be noted that the absorption and emission characteristics of the bound dye can differ from the unbound dye. Therefore, when referring to the various wavelength ranges and characteristics of the dyes, it is intended to indicate the dyes as employed and not the dye which is unconjugated and characterized in an arbitrary solvent.

Fluorescers are generally preferred because by irradiating a fluorescer with light, one can obtain a plurality of emissions. Thus, a single label can provide for a plurality of measurable events.

Detectable signal can also be provided by chemiluminescent and bioluminescent sources. Chemiluminescent sources include a compound which becomes electronically excited by a chemical reaction and can then emit light which serves as the detectable signal or donates energy to a fluorescent acceptor. Alternatively, luciferins can be used in conjunction with luciferase or lucigenins to provide bioluminescence. Spin labels are provided by reporter molecules with an unpaired electron spin which can be detected by electron spin resonance (ESR) spectroscopy. Exemplary spin labels include organic free radicals, transitional metal complexes, particularly vanadium, copper, iron, and manganese, and the like. Exemplary spin labels include nitroxide free radicals.

The label may be added to the target (sample) nucleic acid(s) prior to, or after the hybridization. So called "direct labels" are detectable labels that are directly attached to or incorporated into the target (sample) nucleic acid prior to hybridization. In contrast, so called "indirect labels" are joined to the hybrid duplex after hybridization. Often, the indirect label is attached to a binding moiety that has been attached to the

10

15

20

25

30

target nucleic acid prior to the hybridization. Thus, for example, the target nucleic acid may be biotinylated before the hybridization. After hybridization, an avidin-conjugated fluorophore will bind the biotin bearing hybrid duplexes providing a label that is easily detected. For a detailed review of methods of labeling nucleic acids and detecting labeled hybridized nucleic acids see Laboratory Techniques in Biochemistry and Molecular Biology, Vol. 24: Hybridization With Nucleic Acid Probes, P. Tijssen, ed. Elsevier, N.Y., (1993)).

Fluorescent labels are easily added during an in vitro transcription reaction. Thus, for example, fluorescein labeled UTP and CTP can be incorporated into the RNA produced in an *in vitro* transcription.

The labels can be attached directly or through a linker moiety. In general, the site of label or linker-label attachment is not limited to any specific position. For example, a label may be attached to a nucleoside, nucleotide, or analogue thereof at any position that does not interfere with detection or hybridization as desired. For example, certain Label-ON Reagents from Clontech (Palo Alto, CA) provide for labeling interspersed throughout the phosphate backbone of an oligonucleotide and for terminal labeling at the 3' and 5' ends. As shown for example herein, labels can be attached at positions on the ribose ring or the ribose can be modified and even eliminated as desired. The base moieties of useful labeling reagents can include those that are naturally occurring or modified in a manner that does not interfere with the purpose to which they are put. Modified bases include but are not limited to 7-deaza A and G, 7-deaza-8-aza A and G, and other heterocyclic moieties.

It will be recognized that fluorescent labels are not to be limited to single species organic molecules, but include inorganic molecules, multi-molecular mixtures of organic and/or inorganic molecules, crystals, heteropolymers, and the like. Thus, for example, CdSe-CdS core-shell nanocrystals enclosed in a silica shell can be easily derivatized for coupling to a biological molecule (Bruchez et al. (1998) Science, 281: 2013-2016). Similarly, highly fluorescent quantum dots (zinc sulfide-capped cadmium selenide) have been covalently coupled to biomolecules for use in ultrasensitive biological detection (Warren and Nie (1998) Science, 281: 2016-2018).

Amplification-based assays.

In another embodiment, amplification-based assays can be used to detect nucleic acids. In such amplification-based assays, the nucleic acid sequences act as a

10

15

20

25

30

template in an amplification reaction (e.g. Polymerase Chain Reaction (PCR). Detailed protocols for quantitative PCR are provided in Innis et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, Inc. N.Y.).

Other suitable amplification methods include, but are not limited to ligase chain reaction (LCR) (see Wu and Wallace (1989) Genomics 4: 560, Landegren et al. (1988) Science 241: 1077, and Barringer et al. (1990) Gene 89: 117, transcription amplification (Kwoh et al. (1989) Proc. Natl. Acad. Sci. USA 86: 1173), and self-sustained sequence replication (Guatelli et al. (1990) Proc. Nat. Acad. Sci. USA 87: 1874).

Detection of C. pneumoniae gene expression

The nucleic acids of the invention can also be used to *C. pneumoniae* detect gene transcripts. Methods of detecting and/or quantifying gene transcripts using nucleic acid hybridization techniques are known to those of skill in the art (see Sambrook *et al. supra*). For example, a Northern transfer may be used for the detection of the desired mRNA directly. In brief, the mRNA is isolated from a given cell sample using, for example, an acid guanidinium-phenol-chloroform extraction method. The mRNA is then electrophoresed to separate the mRNA species and the mRNA is transferred from the gel to a nitrocellulose membrane. As with the Southern blots, labeled probes are used to identify and/or quantify the target mRNA.

In another preferred embodiment, the gene transcript can be measured using amplification (e.g. PCR) based methods as described above for directly assessing copy number of the target sequences.

Expression of C. pneumoniae proteins

The nucleic acids disclosed here can be used for recombinant expression of the proteins. In these methods, the nucleic acids encoding the proteins of interest are introduced into suitable host cells, followed by induction of the cells to produce large amounts of the protein. The invention relies on routine techniques in the field of recombinant genetics, well known to those of ordinary skill in the art. A basic text disclosing the general methods of use in this invention is Sambrook *et al.*, *Molecular Cloning*, *A Laboratory Manual* (2nd ed. 1989).

Standard transfection methods are used to produce prokaryotic, mammalian, yeast or insect cell lines which express large quantities of the desired

10

15

20

25

30

polypeptide, which is then purified using standard techniques (see, e.g., Colley et al., J. Biol. Chem. 264:17619-17622, 1989; Guide to Protein Purification, supra).

The nucleotide sequences used to transfect the host cells can be modified to yield *Chlamydia* polypeptides with a variety of desired properties. For example, the polypeptides can vary from the naturally-occurring sequence at the primary structure level by amino acid, insertions, substitutions, deletions, and the like. These modifications can be used in a number of combinations to produce the final modified protein chain.

The amino acid sequence variants can be prepared with various objectives in mind, including facilitating purification and preparation of the recombinant polypeptide. The modified polypeptides are also useful for modifying plasma half life, improving therapeutic efficacy, and lessening the severity or occurrence of side effects during therapeutic use. The amino acid sequence variants are usually predetermined variants not found in nature but exhibit the same immunogenic activity as naturally occurring protein. In general, modifications of the sequences encoding the polypeptides may be readily accomplished by a variety of well-known techniques, such as site-directed mutagenesis (see Gillman & Smith, Gene 8:81-97 (1979); Roberts et al., Nature 328:731-734 (1987)). One of ordinary skill will appreciate that the effect of many mutations is difficult to predict. Thus, most modifications are evaluated by routine screening in a suitable assay for the desired characteristic. For instance, the effect of various modifications on the ability of the polypeptide to elicit a protective immune response can be easily determined using in vitro assays. For instance, the polypeptides can be tested for their ability to induce lymphoproliferation, T cell cytotoxicity, or cytokine production using standard techniques.

The particular procedure used to introduce the genetic material into the host cell for expression of the polypeptide is not particularly critical. Any of the well known procedures for introducing foreign nucleotide sequences into host cells may be used. These include the use of calcium phosphate transfection, spheroplasts, electroporation, liposomes, microinjection, plasmid vectors, viral vectors and any of the other well known methods for introducing cloned genomic DNA, cDNA, synthetic DNA or other foreign genetic material into a host cell (see Sambrook et al., supra). It is only necessary that the particular procedure utilized be capable of successfully introducing at least one gene into the host cell which is capable of expressing the gene.

10

15

20

25

30

Any of a number of well known cells and cell lines can be used to express the polypeptides of the invention. For instance, prokaryotic cells such as *E. coli* can be used. Eukaryotic cells include, yeast, Chinese hamster ovary (CHO) cells, COS cells, and insect cells.

The particular vector used to transport the genetic information into the cell is also not particularly critical. Any of the conventional vectors used for expression of recombinant proteins in prokaryotic and eukaryotic cells may be used. Expression vectors for mammalian cells typically contain regulatory elements from eukaryotic viruses.

The expression vector typically contains a transcription unit or expression cassette that contains all the elements required for the expression of the polypeptide DNA in the host cells. A typical expression cassette contains a promoter operably linked to the DNA sequence encoding a polypeptide and signals required for efficient polyadenylation of the transcript. The term "operably linked" as used herein refers to linkage of a promoter upstream from a DNA sequence such that the promoter mediates transcription of the DNA sequence. The promoter is preferably positioned about the same distance from the heterologous transcription start site as it is from the transcription start site in its natural setting. As is known in the art, however, some variation in this distance can be accommodated without loss of promoter function.

Following the growth of the recombinant cells and expression of the polypeptide, the culture medium is harvested for purification of the secreted protein. The media are typically clarified by centrifugation or filtration to remove cells and cell debris and the proteins are concentrated by adsorption to any suitable resin or by use of ammonium sulfate fractionation, polyethylene glycol precipitation, or by ultrafiltration. Other routine means known in the art may be equally suitable. Further purification of the polypeptide can be accomplished by standard techniques, for example, affinity chromatography, ion exchange chromatography, sizing chromatography, His₆ tagging and Ni-agarose chromatography (as described in Dobeli *et al.*, *Mol. and Biochem. Parasit.* 41:259-268 (1990)), or other protein purification techniques to obtain homogeneity. The purified proteins are then used to produce pharmaceutical compositions, as described below.

An alternative method of preparing recombinant polypeptides useful as vaccines involves the use of recombinant viruses (e.g., vaccinia). Vaccinia virus is grown

. 5

10

15

20

25

30

in suitable cultured mammalian cells such as the HeLa S3 spinner cells, as described by Mackett et al., in DNA cloning Vol. II: A practical approach, pp. 191-211 (Glover, ed.).

Antibody Production

The proteins of the present invention can be used to produce antibodies specifically reactive with *C pneumoniae* antigens. If isolated proteins are used, they may be recombinantly produced or isolated from *Chlamydia* cultures. Synthetic peptides made using the protein sequences may also be used.

Methods of production of polyclonal antibodies are known to those of skill in the art. In brief, an immunogen, preferably a purified protein, is mixed with an adjuvant and animals are immunized. When appropriately high titers of antibody to the immunogen are obtained, blood is collected from the animal and antisera is prepared. Further fractionation of the antisera to enrich for antibodies reactive to *Chlamydia* proteins can be done if desired (see Harlow & Lane, *Antibodies: A Laboratory Manual* (1988)).

Polyclonal antisera are used to identify and characterize *Chlamydia* in the tissues of patients using, for instance, *in situ* techniques and immunoperoxidase test procedures described in Anderson *et al. JAVMA* 198:241 (1991) and Barr *et al. Vet. Pathol.* 28:110-116 (1991).

Monoclonal antibodies may be obtained by various techniques familiar to those skilled in the art. Briefly, spleen cells from an animal immunized with a desired antigen are immortalized, commonly by fusion with a myeloma cell (see Kohler & Milstein, Eur. J. Immunol. 6:511-519 (1976)). Alternative methods of immortalization include transformation with Epstein Barr Virus, oncogenes, or retroviruses, or other methods well known in the art. Colonies arising from single immortalized cells are screened for production of antibodies of the desired specificity and affinity for the antigen, and yield of the monoclonal antibodies produced by such cells may be enhanced by various techniques, including injection into the peritoneal cavity of a vertebrate host.

Monoclonal antibodies produced in such a manner are used, for instance, in ELISA diagnostic tests, immunoperoxidase tests, immunohistochemical tests, for the *in vitro* evaluation of spirochete invasion, to select candidate antigens for vaccine development, protein isolation, and for screening genomic and cDNA libraries to select appropriate gene sequences.

10

15

20

25

30

Immunodiagonostic detection of C. pneumoniae infections

The present invention also provides methods for detecting the presence or absence of *C. pneumoniae*, or antibodies reactive with it, in a biological sample. For instance, antibodies specifically reactive with *Chlamydia* can be detected *using* either *Chlamydia* proteins or the isolates described here. The proteins and isolates can also be used to raise specific antibodies (either monoclonal or polyclonal) to detect the antigen in a sample. In addition, the nucleic acids disclosed and claimed here can be used to detect *Chlamydia*-specific sequences using standard hybridization techniques.

For a review of immunological and immunoassay procedures in general, see *Basic and Clinical Immunology* (Stites & Terr ed., 7th ed. 1991)). The immunoassays of the present invention can be performed in any of several configurations, which are reviewed extensively in *Enzyme Immunoassay* (Maggio, ed., 1980); Tijssen, *Laboratory Techniques in Biochem.stry and Molecular Biology* (1985)). For instance, the proteins and antibodies disclose I here are conveniently used in ELISA, immunoblot analysis and agglutination assays.

In brief, immunoassays to measure anti-Chlamydia antibodies or antigens can be either competitive or noncompetitive binding assays. In competitive binding assays, the sample analyte (e.g., anti-Chlamydia antibodies) competes with a labeled analyte (e.g., anti-Chlamydia monoclonal antibody) for specific binding sites on a capture agent (e.g., isolated Chlamydia protein) bound to a solid surface. The concentration of labeled analyte bound to the capture agent is inversely proportional to the amount of free analyte present in the sample.

Noncompetitive assays are typically sandwich assays, in which the sample analyte is bound between two analyte-specific binding reagents. One of the binding agents is used as a capture agent and is bound to a solid surface. The second binding agent is labelled and is used to measure or detect the resultant complex by visual or instrument means.

A number of combinations of capture agent and labelled binding agent can be used. For instance, an isolated *Chlamydia* protein or culture can be used as the capture agent and labelled anti-human antibodies specific for the constant region of human antibodies can be used as the labelled binding agent. Goat, sheep and other non-luman antibodies specific for human immunoglobulin constant regions (e.g., γ or μ) are

10

15

20

25

30

well known in the art. Alternatively, the anti-human antibodies can be the capture agent and the antigen can be labelled.

Various components of the assay, including the antigen, anti-Chlamydia antibody, or anti-human antibody, may be bound to a solid surface. Many methods for immobilizing biomolecules to a variety of solid surfaces are known in the art. For instance, the solid surface may be a membrane (e.g., nitrocellulose), a microtiter dish (e.g., PVC or polystyrene) or a bead. The desired component may be covalently bound or noncovalently attached through nonspecific bonding.

Alternatively, the immunoassay may be carried out in liquid phase and a variety of separation methods may be employed to separate the bound labeled component from the unbound labelled components. These methods are known to those of skill in the art and include immunoprecipitation, column chromatography, adsorption, addition of magnetizable particles coated with a binding agent and other similar procedures.

An immunoassay may also be carried out in liquid phase without a separation procedure. Various homogeneous immunoassay methods are now being applied to immunoassays for protein analytes. In these methods, the binding of the binding agent to the analyte causes a change in the signal emitted by the label, so that binding may be measured without separating the bound from the unbound labelled component.

Western blot (immunoblot) analysis can also be used to detect the presence of antibodies to *Chlamydia* in the sample. This technique is a reliable method for confirming the presence of antibodies against a particular protein in the sample. The technique generally comprises separating proteins by gel electrophoresis on the basis of molecular weight, transferring the separated proteins to a suitable solid support, (such as a nitrocellulose filter, a nylon filter, or derivatized nylon filter), and incubating the sample with the separated proteins. This causes specific target antibodies present in the sample to bind their respective proteins. Target antibodies are then detected using labeled antihuman antibodies.

The immunoassay formats described above employ labelled assay components. The label may be coupled directly or indirectly to the desired component of the assay according to methods well known in the art. A wide variety of labels may be used. The component may be labelled by any one of several methods. Traditionally a radioactive label incorporating ³H, ¹²⁵I, ³⁵S, ¹⁴C, or ³²P was used. Non-radioactive labels

10

15

20

25

30

include ligands which bind to labelled antibodies, fluorophores, chemiluminescent agents, enzymes, and antibodies which can serve as specific binding pair members for a labelled ligand. The choice of label depends on sensitivity required, ease of conjugation with the compound, stability requirements, and available instrumentation.

Enzymes of interest as labels will primarily be hydrolases, particularly phosphatases, esterases and glycosidases, or oxidoreductases, particularly peroxidases. Fluorescent compounds include fluorescein and its derivatives, rhodamine and its derivatives, dansyl, umbelliferone, etc. Chemiluminescent compounds include luciferin, and 2,3-dihydrophthalazinediones, e.g., luminol. For a review of various labelling or signal producing systems which may be used, see U.S. Patent No. 4,391,904, which is incorporated herein by reference.

Non-radioactive labels are often attached by indirect means. Generally, a ligand molecule (e.g., biotin) is covalently bound to the molecule. The ligand then binds to an anti-ligand (e.g., streptavidin) molecule which is either inherently detectable or covalently bound to a signal system, such as a detectable enzyme, a fluorescent compound, or a chemiluminescent compound. A number of ligands and anti-ligands can be used. Where a ligand has a natural anti-ligand, for example, biotin, thyroxine, and cortisol, it can be used in conjunction with the labelled, naturally occurring anti-ligands. Alternatively, any haptenic or antigenic compound can be used in combination with an antibody.

Some assay formats do not require the use of labelled components. For instance, agglutination assays can be used to detect the presence of the target antibodies. In this case, antigen-coated particles are agglutinated by samples comprising the target antibodies. In this format, none of the components need be labelled and the presence of the target antibody is detected by simple visual inspection.

Pharmaceutical Compositions

The peptides or antibodies (typically monoclonal antibodies) of the present invention and pharmaceutical compositions thereof are useful for administration to mammals, particularly humans, to treat and/or prevent *Chlamydia* infections. Suitable formulations are found in *Remington's Pharmaceutical Sciences*, Mack Publishing Company, Philadelphia, PA, 17th ed. (1985).

10

15

20

25

30

The immunogenic peptides or antibodies of the invention are administered prophylactically or to an individual already suffering from the disease. The peptide compositions are administered to a patient in an amount sufficient to elicit an effective immune response to *Chlamydia*. An effective immune response is one that inhibits infection. An amount adequate to accomplish this is defined as "therapeutically effective dose" or "immunogenically effective dose." Amounts effective for this use will depend on, e.g., the peptide composition, the manner of administration, the stage and severity of the disease being treated, the weight and general state of health of the patient, and the judgment of the prescribing physician, but generally range for the initial immunization (that is for therapeutic or prophylactic administration) from about 0.1 mg to about 1.0 mg per 70 kilogram patient, more commonly from about 0.5 mg to about 0.75 mg per 70 kg of body weight. Boosting dosages are typically from about 0.1 mg to about 0.5 mg of peptide using a boosting regimen over weeks to months depending upon the patient's response and condition. A suitable protocol would include injection at time 0, 4, 2, 6, 10 and 14 weeks, followed by further booster injections at 24 and 28 weeks.

For therapeutic use, administration should begin at the first sign of infection. This is followed by boosting doses until at least symptoms are substantially abated and for a period thereafter. In some circumstances, loading doses followed by boosting doses may be required. The resulting immune response helps to cure or at least partially arrest symptoms and/or complications. Vaccine compositions containing the peptides are administered prophylactically to a patient susceptible to or otherwise at risk of the infection.

The pharmaceutical compositions (containing either peptides or antibodies) are intended for parenteral or oral administration. Preferably, the pharmaceutical compositions are administered parenterally, e.g., subcutaneously, intradermally, or intramuscularly. Thus, the invention provides compositions for parenteral administration which comprise a solution of the immunogenic polypeptides dissolved or suspended in an acceptable carrier, preferably an aqueous carrier. A variety of aqueous carriers may be used, e.g., water, buffered water, 0.4% saline, 0.3% glycine, hyaluronic acid and the like. These compositions may be sterilized by conventional, well known sterilization techniques, or may be sterile filtered. The resulting aqueous solutions may be packaged for use as is, or lyophilized, the lyophilized preparation being combined with a sterile solution prior to administration. The compositions may contain

10

. 15

20

25

30

pharmaceutically acceptable auxiliary substances as required to approximate physiological conditions, such as buffering agents, tonicity adjusting agents, wetting agents and the like, for example, sodium acetate, sodium lactate, sodium chloride, potassium chloride, calcium chloride, sorbitan monolaurate, triethanolamine oleate, etc.

The compositions may also comprise carriers to enhance the immune response. Useful carriers are well known in the art, and include, e.g., KLH, thyroglobulin, albumins such as human serum albumin, tetanus toxoid, polyamino acids such as poly(lysine:glutamic acid), influenza, hepatitis B virus core protein, hepatitis B virus recombinant vaccine and the like.

For solid compositions, conventional nontoxic solid carriers may be used which include, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharin, talcum, cellulose, glucose, sucrose, magnesium carbonate, and the like. For oral administration, a pharmaceutically acceptable nontoxic composition is formed by incorporating any of the normally employed excipients, such as those carriers previously listed, and generally 10-95% of active ingredient, that is, one or more peptides of the invention, and more preferably at a concentration of 25%-75%.

As noted above, the peptide compositions are intended to induce an immune response to *Chlamydia*. Thus, compositions and methods of administration suitable for maximizing the immune response are preferred. For instance, peptides may be introduced into a host, including humans, linked to a carrier or as a homopolymer or heteropolymer of active peptide units from various *Chlamydia* proteins disclosed here. Alternatively, a "cocktail" of polypeptides can be used. A mixture of more than one polypeptide has the advantage of increased immunological reaction and, where different peptides are used to make up the polymer, the additional ability to induce antibodies to a number of epitopes.

The compositions also include an adjuvant. As used here, number of adjuvants are well known to one skilled in the art. Suitable adjuvants include incomplete Freund's adjuvant, alum, aluminum phosphate, aluminum hydroxide, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP),

N-acetyl-nor-muramyl-L-alanyl-D-isoglutamine (CGP 11637, referred to as nor-MDP), N-acetylmuramyl-Lalanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-sn-glycero-3-hydroxyphosphoryloxy)-ethylamine (CGP 19835A, referred to as MTP-PE), and RIBI, which contains three components extracted from bacteria, monophosphoryl

10

15

20

25

30

lipid A, trehalose dimycolate and cell wall skeleton (MPL+TDM+CWS) in a 2% squalene/Tween 80 emulsion. The effectiveness of an adjuvant may be determined by measuring the amount of antibodies directed against the immunogenic peptide.

The concentration of immunogenic peptides of the invention in the pharmaceutical formulations can vary widely, i.e. from less than about 0.1%, usually at or at least about 2% to as much as 20% to 50% or more by weight, and will be selected primarily by fluid volumes, viscosities, etc., in accordance with the particular mode of administration selected.

The peptides of the invention can also be expressed by attenuated viral hosts, such as vaccinia or fowlpox. This approach involves the use of vaccinia virus as a vector to express nucleotide sequences that encode the peptides of the invention. Upon introduction into a host, the recombinant vaccinia virus expresses the immunogenic peptide, and thereby elicits an immune response. Vaccinia vectors and methods useful in immunization protocols are described in, e.g., U.S. Patent No. 4,722,848. Another vector is BCG (Bacille Calmette Guerin). BCG vectors are described in Stover et al. (*Nature* 351:456-460 (1991)). A wide variety of other vectors useful for therapeutic administration or immunization of the peptides of the invention, e.g., *Salmonella typhi* vectors and the like, will be apparent to those skilled in the art from the description herein.

The DNA encoding one or more of the peptides of the invention can also be administered to the patient. This approach is described, for instance, in Wolff et. al., Science 247: 1465-1468 (1990) as well as U.S. Patent Nos. 5,580,859 and 5,589,466.

In order to enhance serum half-life, the peptides may also be encapsulated, introduced into the lumen of liposomes, prepared as a colloid, or other conventional techniques may be employed which provide an extended serum half-life of the peptides. A variety of methods are available for preparing liposomes, as described in, e.g., Szoka et al., *Ann. Rev. Biophys. Bioeng.* 9:467 (1980), U.S. Pat. Nos. 4, 235,871, 4,501,728 and 4,837,028.

EXAMPLES

The following examples are offered to illustrate, but no to limit the claimed invention.

Example 1:

30

10

15

20

25

30

This example describes comparison of the *C. pneumoniae* genome disclosed here and the, previously sequenced, *C. trachomatis* genome (Stephens, *et al. Science* **282:**754-759 (1998)).

The apparent low level of DNA homology between C. trachomatis and C.

pneumoniae (Campbell, et al., J. Clin. Microbiol. 25:1911-1916 (1987)) yet analogous cell structures and developmental cycles, predicts that comparative analysis of the two genomes will significantly enhance the understanding of both pathogens. Identification of genes that are present in one species but not the other are of particular importance for the mutually exclusive biological, virulence and pathogenesis capabilities of each. Identification of genes shared between the two species strongly supports the requirement for these capabilities in a biological system that has, over its long-term association with mammalian host cells, evolved to reduce the metabolic capacities while optimizing survival, growth and transmission of these unique pathogens.

The previously sequenced *C. trachomatis* genome contains 1,042,519 nucleotides and 875 likely protein-coding genes. Similarity searching permitted the inferred functional assignment of sequences 636 (60%) genes disclosed here and 251 (23%) are similar to hypothetical genes for other bacterial organisms including those for *C. trachomatis*. The remaining 186 (17%) genes are not homologous to sequences deposited in GenBank.. Seventy *C. trachomatis* genes are not represented in the *C. pneumoniae* genome. These are contained within blocks consisting of 2-17 genes and 19 single genes. Of the 70 *C. trachomatis* genes without homologs in *C. pneumoniae*, 60 are classified as encoding hypothetical proteins. The remaining genes not represented in *C. pneumoniae* consist of the tryptophan operon (*trpA,B,R*), *trpC*, two predicted thiol protease genes, and 4 genes assigned to the phospholipase-D superfamily.

It is evident that there is a high level of functional conservation between C. pneumoniae and C. trachomatis as orthologs to C. trachomatis genes were identified for 859 (80%) of the predicted coding sequences for C. pneumoniae. The level of similarity for individual encoded proteins spans a wide spectrum (22-95% amino acid identity) with an average of 62% amino acid identity between orthologs from the two species. The percent amino acid identity between orthologous chlamydial proteins is similar among functional groups with the highest for proteins associated with translation and the lowest for proteins whose function in chlamydiae is uncharacterized and not related to proteins encoded by other organisms. The gene order of the homologous set of genes in C.

10

15

20

25

30

pneumoniae shows reorganization relative to the genome of *C. trachomatis*; however, there is a high level of synteny for the gene organization of the two genomes. We identified thirty-nine blocks of 2 or more genes whose gene organization is colinear with homologs to *C. trachomatis*, although some of these are inverted. The distribution of genome reorganization is not evenly distributed on the chromosome as the region between *C. pneumoniae* coding sequences 0130-0300 contains substantially more reorganization than other areas of the genome. This region coincides with the predicted chromosome replication terminus.

We identified orthologs of enzymes characterized in other bacteria that account for the essential requirements for DNA replication, repair, transcription and translation including two predicted DNA helicases of the Swi2/Snf2 family found in *C. trachomatis*. Similar to *C. trachomatis*, alternative sigma subunits for RNA polymerase, σ²⁸ and σ⁵⁴, were identified in addition to anti-σ regulatory system factors RsbV, a RsbW-like single-domain histidine kinase, and a RsbU-like protein phosphatase. These findings suggest that the fundamental mechanisms of transcriptional regulation are conserved among *Chlamydia*. The *C. trachomatis* proteins containing SET and SWIB domains, and a SWIB domain fused to the C-terminus of the chlamydial topoisomerase I, not identified outside eukaryotes, are found in *C. pneumoniae* supporting their possible role in the chromatin condensation-decondensation characteristic of the biologically unique chlamydial developmental cycle.

The central metabolic pathways inferred from the *C. pneumoniae* genome sequence are the same as those identified for *C. trachomatis C. pneumoniae* has a glycolytic pathway and a linked tricarboxylic acid cycle, although likely functional, is incomplete as genes for citrate synthase, aconitase, and isocitrate dehydrogenase were not identified. *C. pneumoniae* has a complete glycogen synthesis and degradation system supporting a role for glycogen synthesis and utilization of glucose-derivatives in chlamydial metabolism. Genes encoding essential functions in aerobic respiration are present and electron flux may be supported by pyruvate, succinate, glycerol-3-phosphate, and NADH dehydrogenases, NADH-ubiquinone oxidoreductase and cytochrome oxidase.

C. pneumoniae also contains the V (vacuolar)-type ATPase operon and the two ATP translocases found in C. trachomatis.

The type-III secretion virulence system required for invasion by several pathogenic bacteria and found in the *C. trachomatis* genome in three chromosomal

10

15

20

25

30

locationsis also present in the *C. pneumoniae* genome. Each of the components is conserved and their relative genomic contexts are conserved. Genes such as a predicted serine/threonine protein kinase and other genes physically linked to genes encoding structural components of the type-III secretion apparatus, but without identified homologs, are also highly similar between the two species suggesting the functional roles in modifying cellular biology are fundamentally conserved.

Chlamydia-encoded proteins that are not found in chlamydial organisms but localized to the intracellular chlamydial inclusion membrane are likely essential for the unique intracellular biology and perhaps differences in inclusion morphology observed between species of Chlamydia. Several such proteins, termed IncA,B&C, have been characterized for a C. psittaci strain (Rockey, et al. Mol. Microbiol. 15:617-626 (1995); Rockey et al. Infect. Immun. 62:106-112 (1994)). C. pneumoniae and C. trachomatis encode orthologs to C. psittaci IncB and IncC and C. trachomatis also contains an ortholog to IncA. C. pneumoniae contains two genes that encode proteins with similarity to IncA (CPn0186 and CPn0585), although the level of homology is low suggesting analogous but possibily altered functions.

The tryptophan biosynthesis operon (trpA, trpB, trpR) and trpC identified in C. trachomatis is conspicuously missing in the C. pneumoniae genome. This represents the entire repertoire of genes associated with tryptophan biosynthesis identified in C. trachomatis. Seventeen genes adjacent to the C. trachomatis tryptophan operon also were not found in the C. pneumoniae genome. This region is the single largest loss of a contiguous genomic segment and includes 4 HKD superfamily encoding genes that encompass a family of proteins related to endonuclease and phospholipase D. These findings may be important for the ability of Chlamydia to persist in their hosts and cause disease by eliciting potent, focal and persistent inflammatory responses thought to be essential for pathogenesis.

The *C. pneumoniae* genome contains 187,711 additional nucleotides compared to the *C. trachomatis* genome, and the 214 coding sequences not found in *C. trachomatis* account for most of the increased genome size. Eighty-eight of these genes are found in blocks of >10 genes (11-30 genes/block), 41 are single genes, and the remainder are partnered with at least one other gene. Based upon the observation that ~70% of all the *C. pneumoniae* genes have an identifiable homolog in GenBank, exclusive of *C. trachomatis*, it would be expected that over 150 of the 214 genes should

10

15

20

25

30

have a homolog in GenBank, many associated with a function. However, only 28 coding sequences have similarity to genes from other organisms. Thus the majority of the genes that are mutually exclusive of C. trachomatis (186 of 214), and the 60 of 70 C. trachomatis genes that lacked an identifiable homolog in C. pneumoniae, do not have detectable homologs to genes from other organisms. We predict that most of the unique genes are essential for specific attributes that define the differential biology, tropism and pathogenesis of C. trachomatis and C. pneumoniae. Moreover, this suggests that C. pneumoniae has more unique biological (i.e., virulence) capacity than C. trachomatis. The ability of C. pneumoniae to be more invasive and survive in a broader range of host cell types than C. trachomatis is consistent with this hypothesis. Not all of the differences in biological capacity may be associated with mutually exclusive genes. One explanation for the significantly lower level of homology between protein sequences assigned as having C. pneumoniae and C. trachomatis orthologs but no identifiable orthologs in other organisms is that this set of proteins is not only associated with biological requirements specific for Chlamydia but this polymorphism may account for differential biology between the two species. The determination of the genome sequence from a representative of the C. psittaci group will precisely delineate those genes that are mutually exclusive and specific for each species.

The major functionally identifiable addition to the *C. pneumoniae* genome is a large expansion of genes encoding a new family of chlamydial polymorphic membrane proteins (Pmp), alone representing 22% of the increased coding capacity. While the *C. trachomatis* genome has 9 *pmp* genes, remarkably the *C. pneumoniae* genome contains 21 *pmp* genes. Most of these genes appear to be amplified in two regions of the genome with three stand-alone genes. Interestingly one of the stand-alone genes is most closely related to the *C. trachomatis pmpD* which is the only stand-alone *pmp* gene in the *C. trachomatis* genome and it is located with the same relative genomic context, suggesting an essential and conserved function for this paralog. Six Pmp-coding genes are presumably not functional as five contain predicted coding frame-shifts and one is truncated. The amplification of this gene family and the confidently predicted frame-shifts suggest a specific molecular mechanism to promote functional or antigenic diversity. The biological role of this protein family remains enigmatic, although at least one of the proteins in *C. psittaci* related to this family is exposed on the chlamydial surface.

10

15

20

25

30

While a function could not be assigned for most of the unique *C. pneumoniae* genes, several have significant similarity to genes from other organisms. Functional assignments could be made for genes encoding GMP synthetase, IMP dehydrogenase, UMP synthase, uridine kinase, biotin synthase pathway proteins, methylthioadenosine nucleosidase, a DNA glycosylase and aromatic amino acid hydroxylase. Thus a complete pathway was identified for biotin biosynthesis. The additional purine and pyrimidine salvage pathway genes presumably reflect metabolic limitations in one of the cell types that *C. pneumoniae* infects or differences in the ability of *C. pneumoniae* to transport precursor nucleosides or nucleotides.

The addition of aromatic amino acid hydroxylase in *C. pneumoniae* is intriguing especially in light of the loss of tryptophan biosynthetic genes and the inability to synthesize other amino acids including phenylalanine. Aromatic amino acid hyroxlyases include three distinct enzymes that function to receptively oxidize phenylalanine to tyrosine, tyrosine to Dopa, and tryptophan to 5-hydroxytryptophan and serotonin. Although the chlamydial protein is similar to proteins of this family and incrementally more closely related to tryptophan hydroxylase, its specific function could not be confidently predicted. We hypothesize that it may be involved in *C. pneumoniae* virulence. Tryptophan hydroxylase has not been previously identified in bacteria and the origin of the chlamydial gene appears to be from eukaryotes. The functional role of an aromatic amino acid hydroxylase for *C. pneumoniae* is linked to the unique intracellular biology of this organism and may represent a key contribution to *C. pneumoniae* persistence and pathogenesis.

It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art and are to be included within the spirit and purview of this application and scope of the appended claims. All publications, patents, and patent applications cited herein are hereby incorporated by reference in their entirety for all purposes.

Table 1 provides functional assignments of *C. pneumoniae* nonprotein-encoding genomic sequences. Table 2 provides functional assignments of protein coding sequences. Table 3 provides the amino acid sequences of the proteins corresponding to the coding sequences.

TABLE 1

type	SEQ ID NO:1 start position	SEQ ID NO end positio		
OTRINA OT	start position 841664 138493 607342 1000564 10005393 269070 164318 296224 836191 1030533 784896 781680 961536 999949 268992 672236 680178 715863 784994 843926 409922 631373 677337 807413 877473 462141 1085605 786780 89728 293477 87522 199301 199390 626904 708359 1142034 1230028	end positio 841396 138074 607649 1002115 1005278 10055509 269142 164389 296151 836119 1030603 784822 781610 961607 1000023 269065 672318 680257 715944 784922 843999 409848 631445 677264 807341 877400 462214 1085676 786708 89657 293405 199317 626987 708440 1142117 1229945	(R) Putative Origin of R (R) tmRNA Ribonuclease P RNA 16S rRNA 23S rRNA 5S rRNA Ala tRNA_1 Asn tRNA (R) Asp tRNA (R) Ala tRNA_2 Cys tRNA (R) Glu tRNA (R) Glu tRNA (R) Gly tRNA_1 Gly tRNA_1 Het tRNA Leu tRNA_2 Leu tRNA_3 Leu tRNA_4 Leu tRNA_4 Leu tRNA_5 (R) Lys tRNA Pro tRNA_2 (R) Pro tRNA_1 Phe tRNA (R) Arg tRNA_2 (R) Arg tRNA_3 (R) Arg tRNA_3 (R) Arg tRNA_4 Arg tRNA_1 Gln tRNA (R) Thr tRNA_3 (R) Thr tRNA_1 (R) Thr tRNA_1 (R) Thr tRNA_2 (R) Met tRNA_1 (R) Met tRNA_2 (R) Met tRNA_1 (R) Met tRNA_2 (R) Met tRNA_3 Ser tRNA_1 Ser tRNA_1 Ser tRNA_2 Ser tRNA_3 (R) Ser tRNA_4	leplica
tRNA tRNA tRNA	91070 293399 296147	90999 293317 296075	(R) Trp tRNA (R) Tyr tRNA (R) Val tRNA_1 Val tRNA_2	
tRNA	1137389	1137462	AGT CIMUTE	

TABLE 2

```
Gene function (C. trachomatis ortholog in parentheses)
                               Strand
Gene !
                                        CT001 hypothetical protein
CPn0001
           282
                                   R
                                        gatC-Glu-tRNA Gln Amidotransferase (C subunit) - (CT002)
           573
                       875
CP00002
                                        gatA-Glu tRNA Gln Amidotransferae-(CT003)
                       2370
CPn0003
           895
                                   F
                                        gar8-(Petil2) GlocePHA Gin Amidotransterase (8 Subunit)-(CT904)
                        . ( )
                                   17
 100004
           2.170
                                         pmp_1-Polymorphic Outer Membrane Protein G Family
                       6632
CPN0005
           4127
                                   F
CPn0006
           7293
                       7141
                                   R
                       10496
CPn0007
           7605
CPn0008
           10975
                       11685
                       13119
CPn0009
           11815
CPn0010
           13435
                       14325
                                        frame-shift with 0010
                       15746
           14379
                                   P
CPn0010
                       16614
           15892
CPn0011
                       18212
                                   P
CPn0012
           16644
                                        pmp_2-Polymorphic Outer Membrane Protein G Family
                       21106
                                   F
CPn0013
           18584
                                        pmp_3-Polymorphic Outer Membrane Protein G Family
           21392
                       21922
                                   F
CPn0014
                                        pmp_3-PMP_3 (frame-shift with 0014)
                       24174
                                  Į,
CPn0015
           21835
                                        pmp_4-Polymorphic Outer Membrane Protein G Family
                       26188
CPn0016
           24416
                                  F'
                                        pmp_4-PMP_4 (frame-shift with 0016)
                       27170
           26094
CPn0017
                                        pmp_5-Polymorphic Outer Membrane Protein G Family
CPn0018
           27522
                       29003
                                  Г
                                        pmp_5-PMP_5 (frame-shift with 0018)
           29007
                       30356
                                  F
CPn0019
                                        Predicted OMP [leader (14) peptide: outer membrane]-(CT351)
CP0020
           32687
                       30603
                                  P.
                                        Predicted OMP [leader (19) peptide]-(CT350)
                       32707
CPn0021
           34410
                                  R
CP50022
           34982
                       34395
                                        maf-(CT349)
                                        yjjK/alr-ABC Transporter Protein ATPase-(CT348)
                       35014
                                  R
CP#0023
           36603
                                        xerC-Integrase/recombinase-(CT347)
CPn0024
           37596
                       36661
                                        elaC/atsA-Sulphohydrolase/Glycosulfatase-( CT346)
           38604
                       37684
                                  R
CP00025
                                        CT345 hypothetical protein-(CT345)
           39625
                       38762
                                  R
CPn0026
                                        lon-Lon ATP-dependent Protease-(CT344)
                       39778
                                  R
CPn0027
           42234
                       42543
                                  R
CPp0028
           43325
CPn0029
           43755
                       43390
                                  R
                                        gcp_1-0-Sialoglycoprotein Endopeptidase_1-(CT343)
                       44529
CP0030
           43891
                                  ₽
                                        rs21-S21 Ribosomal Protein-(CT342)
CPn0031
           44711
                       44884
                                        dnaJ-Heat Shock Protein J-(CT341)
                       46098
CP##032
           44923
                                  F
                                        pdhA&B/odbA&odbB-(pyruvate) Oxoisovalerate Dehydrogenase Alpha & Beta
           46138
                       48171
                                  F
CPm@033
                                           Fusion-(CT340)
CPn0034
           49457
                       48210
                                  R
CPh0035
           51029
                       49569
                                        CT339 hypothetical protein
                                  R
           51002
                       51796
                                        CT338 hypothetical protein
CPn0036
                                  F
                                        ptsH-PTS Phosphocarrier Protein Hpr-(CT337)
CPn0037
           51792
                       52115
                                  F
                                        ptsI-PTS PEP Phosphotransferase-(CT336)
                       53831
CPn0038
           52119
                                  F
€Pn0039
           54250
                       53963
                                        ybaB-(CT335)
                                        dnaX_1-DNA Pol III Gamma and Tau_1-(CT334)
CPn0040
           55643
                       54318
                                  R
CPn0041
           55996
                       57342
CPn0042
           57403
                       58182
           58447
                       60372
                                  F
CPn0043
                       60778
CPn0044
           60419
           61069
                       62790
                                  y
CPn0045
CPn0046
           62790
                       63263
                                  F
CPn0047
           63455
                       63652
                                  F
                                        *yqfF-Bs conserved hypothetical IM protein
CPn0048
           63687
                       65801
CPn0049
           66296
                       65817
CPn0050
           66813
                       66499
                                  R
           66833
                       67111
CPn0051
                                        hemC-Porphobilinogen Deaminase-(CT299)
           68005
CPn0052
                       67304
                                  R
                       67986
                                        sms-Sms Protein-(CT298)
CPn0053
           69344
                                        rnc-Ribonuclease III-(CT297)
                       69313
CPn0054
           70023
                                  R
                       70590
                                        CT296 hypothetical protein
CPn0055
           70129
                                  F
                                        mrsA-Phosphomannomutase-(CT295)
           70953
                       72746
CPn0056
                                        sodM-Superoxide Dismutase (Mn) - (CT294)
CPn0057
           72934
                       73554
                                        accD-AcCoA Carboxylase/Transferase Beta-(CT293)
CPn0058
           73639
                       74562
                                        dut-dUTP Nucleotidohydrolase-(CT292)
           74616
                       75050
CPn0059
                                  F
                                        ptsN_1-PTS [[A Protein-(CT291)
                       75528
CPn0060
           75055
                                        ptsN_2-PTS IIA Protein + HTH DNA-Binding Domain-(CT290)
                       76208
CPn0061
           75534
                                  F
                                        CT189 hypothetical protein
           76308
                       77690
CPn0062
                       78267
           78112
CPn0063
                                  F
                       78576
CPn0054
           78346
                                   F
                                        CT288 hypothetical protein
                       30651
CPn0065
           78924
                                   F
           30925
                       82655
C200055
```

```
84053
CPn0067
           82953
                                       CT360 hypothetical protein
CPn0068
           84903
                      84331
                                 R
           85236
                      87086
CPn0069
CPn0070
           87378
                      87208
                                 R
                                       CT325 hypothetical protein
                      87599
                                 R
           88045
CPn0071
                                       CT324 hypothetical protein
CPn0072
           89061
                      88057
                                  R
                                       infA-Initiation Factor IF-1-(CT323)
                      89574
CPn0073
           89356
                                 F
                                       tufA-Elongation Factor Tu-(CT322)
                      90955
CPn0074
           89774
                                 F
                                 F
                                       secE-preprotein translocase-(CT321)
                      91350
CPn0075
           91102
                                       nusG-Transcriptional Antitermination-(CT320)
                      91903
                                 F
CPn0076
           91358
                                       rlll-Lll Ribosomal Protein-(CT319)
           92013
                      92435
CPn0077
                                       rll-L1 Ribosomal Protein-(CT318)
CPn0078
           92465
                      93160
                                 F
                                       rl10-L10 Ribosomal Protein-(CT317)
                      93688
           93179
CPn0079
                                       rl7-L7/L12 Ribosomal Protein-(CT316)
                      94121
CPn0080
           93735
                                       rpoB-RNA Polymerase Beta-(CT315)
CPn0081
           94261
                      98016
                                 F
                                       rpoC-RNA Polymerase Beta' -(CT314)
           98043
                      102221
CPn0082
                                       tal-Transaldolase-(CT313)
           102332
                      103312
                                 F
CPn0083
                      103751
                                       predicted ferredoxin-(CT312)
           103362
CPn0084
                                       CT311 hypothetical protein
           104506
                      103766
                                 R
CPn0085
                                       atpE-ATP Synthase Subunit E-(CT310)
                      105527
                                 F
           104904
CPn0086
                                       CT309 hypothetical protein
CPn0087
           105579
                      106376
                                 F
                                       atpA-ATP Synthase Subunit A-(CT308)
                      108145
                                 F
CPn0088
           106373
                                       atpB-ATP Synthase Subunit B-(CT307)
           108153
                      109466
CPn0089
                                       atpD-ATP Synthase Subunit D-(CT306)
                      110080
           109454
                                 F
C2110090
                                       atpI-ATP Synthase Subunit I-(CT305)
CPn0091
           110074
                      112053
                                 F
                                       atpK-ATP Synthase Subunit K-(CT304)
           112151
                      112573
                                 F
CPn0092
                                       CT303 hypothetical protein
                      113015
           112509
CPn0093
                                       valS-Valyl tRNA Synthetase-(CT302)
CPn0094
           113152
                      115971
                                       pknD-S/T Protein Kinase-(CT301)
CP00095
                      118790
           116037
                                 F
                                       uvrA-Excinuclease ABC Subunit A-(CT333)
C$p.0096
           124314
                      118837
                                 R
                                       pyk-Pyruvate Kinase-(CT332)
CPn0097
           124555
                      126006
                                 F
           127491
                      126091
                                       htrB-Acyltransferase-(CT010)
Chn0098
                                 R
CPn0099
           127593
                      127865
                                 F
                                       CT011 hypothetical protein
CPn0100
           129141
                      127882
                                 R
                                       ybbP family hypothetical protein-(CT012)
CPn0101
           129932
                      129141
                                 R
                                       cydA-Cytochrome Oxidase Subunit I-(CT013)
                      131466
CPR0102
           130123
                                 F
                                       cydB-Cytochrome Oxidase Subunit II-(CT014)
                      132511
CPm0103
           131480
                                       CT017 hypothetical protein
                      132676
           133875
                                 R
CPn0104
CPm0105
           134847
                      134029
                                  R
                                       CT016 hypothetical protein
                                       phoH-ATPase-(CT015)
CP0106
           135091
                      136374
                                 F
                                       CT058 hypothetical protein_1
CPh0107
           137162
                      136392
CPn0108
           137857
                      137303
                                  R
                                       CT018
                      141783
                                  F
                                       ileS-Isoleucyl-tRNA Synthetase-(CT019)
CPn0109
           138655
                                       lepB-Signal Peptidase I-(CT020)
           143734
                      141827
CPn0110
                                  R
           144686
                      143934
                                       CT021 hypothetical protein
CPn0111
                                  R
                                       rl31-L31 Ribosomal Protein-(CT022)
CPn0112
           144767
                      145093
                                       pfrA-Peptide Chain Releasing Factor (RF-1)-(CT023)
           145335
                      146405
                                 F
CPn0113
                                       hemK-A/G specific methylase-(CT024)
CPn0114
           146398
                      147261
                                 F
                                       ffh-Signal Recognition Particle GTPase-(CT025)
CPn0115
           147279
                      148622
                                 F
                                       rs16-S16 Ribosomal Protein-(CT026)
CPn0116
           148616
                      148972
                                       trmD-tRNA (guanine N-1)-Methyltransferase-(CT027)
CPn0117
           148989
                      150071
                                  F
                                       rl19-L19 Ribosomal Protein-(CT028)
CPn0118
           150102
                      150464
                                · F
                                       rnhB_1-Ribonuclease HII_1-(CT029)
CPn0119
           150523
                       151164
                                       gmk-GMP Kinase-(CT030)
CPn0120
           151164
                      151778
                                  F
CPn0121
           151778
                      152068
                                       CT031 hypothetical protein
                                       metG-Methionyl-tRNA Synthetase-(CT032)
CPn0122
           152071
                      153723
                                  F
                                       recD_1-Exodeoxyribonuclease V (Alpha Subunit)_1-(CT033)
CPn0123
           155969
                      153774
                                  R
CPn0124
           156614
                      158068
CPn0125
           158096
                      158605
                                  F
CPn0126
           158809
                      161085
                                  F
                                       ytfF-Cationic Amino Acid Transporter-(CT034)
           162143
                      161130
CPn0127
                                  R
                                       bpl1-Biotin Protein Ligase-(CT035)
           162277
CPn0128
                       163053
                                  F
                                       similarity to CT036
           163717
CPn0129
                       163064
                                  R
CPn0130
           164245
                       163751
                                  R
CPn0131
           164549
                       165580
CPn0132
           165587
                       166561
                                  F
                                       CHLPS hypothetical protein-(CT109)
CPn0133
           167334
                       166564
                                  R
                                       groEL_1-HSP-60_1-(CT110)
CPn0134
           169098
                       167467
                                  R
                                       groES-10KDa Chaperonin-(CT111)
CPn0135
           169448
                       169143
                                  R
CPn0136
           171401
                       169569
                                  R
                                       pepF-Oligopeptidase-(CT112)
CPn0137
                       171502
                                  R
                                       ybgI-ACR family-(CT108)
           172254
                                       hemL-Glutamate-1-semialdehyde-2,1-aminomutase-(CT210)
CPn0138
           174019
                       172700
                                  R
```

```
CPn0139
           174656
                       174093
                                       yggE-(CT210)
CPn0140
           175110
                       174673
                                  R
                                        yqdE-(CT212)
                                        rpiA-Ribose-5-P Isomerase A-(CT213)
           175802
                       175110
                                  R
CPn0141
CPn0142
           176091
                       175816
                                  R
                                        *yxjG_Bs_1 Hypothetical Protein
                                  R
                       176214
CPn0143
           177335
                                        clpB-Clp Protease ATPase-(CT113)
           177963
                       180560
                                  F
CPn0144
                                        CT114 hypothetical protein
           180777
                       182369
                                  F
CPn0145
                      183095
                                  F
           182613
CPn0146
CPn0147
                       183671
           183225
                                        pknl-S/T Protein Kinase-(CT145)
                       185702
                                  F
           183846
CPn0148
           185715
                       187700
                                  F
                                        dnlJ-DNA Ligase-(CT146)
CPn0149
                                        CT147 hypothetical protein
           187834
                      192444
                                  F
CPn0150
                                        mhpA-Monooxygenase-(CT148)
                       192625
                                  R
           194142
CPn0151
                                        CT149 hypothetical protein
                       194318
                                  R
CPn0152
           195265
                                        leuS-Leucyl tRNA Synthetase-(CT209)
                       197892
                                  F
CPn0153
           195433
                                  F
                                        gseA-KDO Transferase-(CT208)
           197892
                       199202
CPn0154
CPn0155
           199691
                       199488
                                  R
           200117
                       199770
                                  R
CPn0156
CPn0157
           200723
                       200298
                                  R
           201430
                       200894
                                  R
CPn0158
CPn0159
           201772
                       201467
                                  R
                                       pfkA_1-Fructose-6-P Phosphotransferase_1-(CT207)
           203791
                       202127
                                  R
CPn0160
                                        predicted acyltransferase family-(CT206)
CPm0161
           204622
                       203798
                                  R
                       204803
                                  R
CPn0162
           205828
           206026
                       206394
                                  F
CPn 0163
           206498
                       206998
                                  F
CPri0164
           206998
                       207582
                                  F
CPn 0165
           207630
                       207962
CPm0166
           208306
                       207977
CPn0167
                                  R
CPn0168
           208641
                       208417
                                  R
CPn0169
           209501
                      208710
                                  R
CPm0170
           211026
                       210025
                                  R
           212435
                       211149
                                  R
                                        *quaA-GMP Synthase
CPn0171
                                        *guaB/impD-Inosine 5'-monophosphase dehydrogenase (C00H-terminal region
           213177
                       212440
                                  R
CPn0172
                                          only)
CP#0173
           213987
                      213715
                                  R
           214257
                       214724
CPm0174
                       215275
                                  F
CPn0175
           214898
CPn0176
           215286
                       216518
                                  F
                                        CT153 hypothetical protein
CPri0177
                       216608
           217459
                                  R
CPri0178
           218052
                       217789
                                  R
                       218056
CPn0179
           218403
                                  R
CPn0180
           218851
                       218355
                                  R
CPn0181
           219175
                       218777
                                  R
                                       accC-Biotin Carboxylase-(CT124)
CPn0182
           220695
                       219334
                                  R
                                        accB-Biotin Carboxyl Carrier Protein-(CT123)
CPn0183
           221195
                       220695
                                  R
                                        efp_1-Elongation Factor P_1-(CT122)
CPn0184
           221775
                       221221
                                  R
                       221765
                                        rpe/araD-Ribulose-P Epimerase-(CT121)
CPn0185
           222451
                                  R
                       224068
                                        *similarity to Cps IncA_1-(CT119)
CPn0186
           222899
                                  F
                                        predicted methylase-(CT133)
CPn0187
           224248
                       225045
                       226400
                                        CT132 hypothetical protein
CPn0188
           225111
                                  F
                                       CT131 homolog-(Possible Transmembrane Protein)
CPn0189
           226400
                       229825
                                  F
CPn0190
           229919
                       231274
                                  F
                                       glnQ-ABC Amino Acid Transporter ATPase-(CT130)
CPn0191
           231991
                       231314
                                  R
                                       glnP-ABC Amino Acid Transporter Permease-(CT129)
                       231984
CPn0192
           232634
                                  R
                                        *argR-Arginine Repressor
                       232686
CPn0193
           233126
                                  R
                                        gcp_2-O-Sialoglycoprotein Endopeptidase_2-(CT197)
CPn0194
           233210
                       234241
                                  F
                                        oppA_1-Oligopeptide Binding Protein_1
CPn0195
           234190
                       235785
                                  F
CPn0196
           235939
                       237519
                                        oppA_2-Oligopeptide Binding Protein_2-(CT198)
                                  F
                                        oppA_3-Oligopeptide Binding Protein_3
CPn0197
           237578
                       238882
                                  F
                                        oppA_4-Oligopeptide Binding Protein_4
                       240746
                                  F
CPn0198
           239169
                                        oppB_1-Oligopeptide Permease_1-(CT199)
                       241983
CPn0199
           241042
                                  F
                                        oppC_1-Oligopeptide Permease_1-(CT200)
                       242868
                                  F
CPn0200
           242017
                                        oppD-Oligopeptide Transport ATPase-(CT201)
CPn0201
           242864
                       243715
                                  F
                                        oppF-Oligopeptide Transport ATPase-(CT202)
CPn0202
           243715
                       244500
                                  F
CPn0203 --- 245008
                       245802
CPn0204
           245817
                       246002
CPn0205
           246133
                       246327
CPn0206
                                        CT203 hypothetical protein
           246409
                       247161
                                        ybhI/sodiTl-Oxoglutarate/Malate Translocator-(CT204)
CPn0207
           247208
                       248617
                                  F
                                  F
                                        pfkA_2-Fructose-6-P Phosphotransferase_2-(CT205)
CPn02-08
           248953
                       250602
```

CPn0209

251036

```
CPn0210
           252384
                      251440
                                  R
CPn0211
           252756
                       252463
                                  R
CPn0212
           254066
                      252888
                                  R
           254342
                      254190
CPn0213
CPn0214
           255657
                      254446
           257015
                      255759
                                  R
CPn0215
                       257174
CPn0216
           257608
                                  R
                      258579
                                       ypdP-(CT140)
CPn0217
           257896
                                  F
           259058
                      258582
                                  R
CPn0218
                                       tgt-Queuine tRNA Ribosyl Transferase-(CT193)
           259357
                      260472
                                  F
CPn0219
                                  F
           260696
                      261238
CPn0220
           261657
                      262064
                                  F
CPn0221
                                       *weak similarity to Bacteriophage CHP1 (Orf4)
                      262842
                                  F
CPn0222
           262504
           262956
                      263333
                                  F
CPn0223
           263435
                      263674
                                  F
CPn0224
                                  F
           263873
                      264541
CPn0225
                      264967
                                  F
CPn0226
           264566
                                       dsbB-Disulfide bond Oxidoreductase-(CT176)
                      265009
CPn0227
           265416
                                  R
CPn0228
                                       dsbG-Disulfide Bond Chaperone-(CT177)
           266110
                      265412
                                  R
                                       CT178 hypothetical protein
CPn0229
           266328
                      267560
                                  F
           268253
                      267576
                                 R
                                       CT179 hypothetical protein
CPn0230
                                       tauB-ABC Transport ATPase (Nitrate/Fe)-(CT180)
           268957
                      268253
CPn0231
                                  R
                                       *similarity to 5'-Methylthioadenosine / S-Adenosylhomocysteine
                      269232
CPn0232
           270122
                                 R
                                          Nucleosidase
 mğ.
           270424
                      270248
CPn0233
                                  R
CPm0234
                                       CT181 hypothetical protein
           271240
                      270548
                                 R
                                       kdsB-deoxyoctulonosic Acid Synthetase-(CT182)
CPn0235
           271416
                      272177
                                 F
                                       pyrG-CTP Synthetase-(CT183)
           272156
                      273766
CPn0236
                                 F
           273762
                      274214
                                       yggF Family-(CT184)
CP10237
                                       zwf-Glucose-6-P Dehyrogenase-(CT185)
CPn0238
           274303
                      275838
                                 F
                                       devB-Glucose-6-P Dehyrogenase (DevB family)-(CT186)
                                 F
CPn0239
           275899
                      276672
CPn0240
           277861
                      276698
                                 R
CPn0241
           279354
                      278203
                                 R
CP#0242
           279918
                      279487
                                 R
CPm0243
           280555
                      280133
                                 R
                                       adk-Adenylate Kinase-(CT128)
CPn0244
           280918
                      281556
                                 F
                                       ydhO-Polysaccharide Hydrolase-Invasin Repeat Family-(CT127)
           281645
CPn0245
                      282499
                                 F
CPn 0246
           282952
                      282551
                                       rs9-S9 Ribosomal Protein-(CT126)
                                       rl13-L13 Ribosomal Protein-(CT125)
CPG 0247
           283415
                      282969
                                 R
CPn0248
           284327
                      283650
                                 R
                                       ycfV/ybbA-ABC Transporter ATPase-(CT152)
                                       CT151 hypothetical protein
CPn0249
           285841
                      284333
                                 R
CPn0250
           286057
                      285902
                                 R
                                       rl33-L33 Ribosomal Protein-(CT150)
CPn0251
           286060
                      287559
                                 F
                                       *conserved hypothetical protein
CPn0252
           288112
                      287576
                                 R
                                       CT144 hypothetical protein (frame-shift with 0253?)
CPn0253
           288456
                      287950
                                 R
                                       CT144 hypothetical protein_1
                                       CT143 hypothetical protein_1
CPn0254
           289262
                      288459
                                 R
CPn0255
           290165
                                       CT142 hypothetical protein_1
                      289329
                                 R
CPn0256
           291264
                      290398
                                       CT144 hypothetical protein_2
                                 R
CPn0257
           292127
                      291267
                                 R
                                       CT143 hypothetical protein_2
                                       CT142 hypothetical protein (frame-shift with 0259?)
CPn0258
           292534
                      292133
                                 R
CPn0259
           292986
                      292441
                                       CT142 hypothetical protein_2
                                 R
CPn0260
           294045
                      293548
                                       secA_1-Protein Translocase Subunit_1-(CT141)
                                 R
                                       ydaO-PP-Loop Superfamily ATPase-(CT217)
CPn0261
           294302
                      295033
                                 F
CPn0262
           295091
                      295933
                                       surE-SurE-like Acid Phosphatase-(CT218)
                                 F
CPn0263
           296249
                                       yqfU hypothetical protein-(CT221)
                      297136
                                 F
CPn0264
           297730
                      297155
                                       ubiD-Phenylacrylate Decarboxylase-(CT220)
                                 R
CPn0265
                                       ubiA-Benzoate Octaphenyltransferase-(CT219)
           298620
                      297730
                                 R
CPn0266
           299184
                      299876
                                 F
CPn0267
           300122
                      300910
                                 F
CPn0268
           300935
                      301318
                                 F
CPn0269
           302450
                      301476
                                       Dipeptidase-(CT138)
                                 R
                                       ywlC-SuA5 Superfamily-related Protein-(CT137)
CPn0270
           303325
                      302468
                                 R
CPn0271
                                       Lysophospholipase esterase-(CT136)
           303634
                      304362
                                 F
CPn0272
                                       dnaX_2-DNA Pol III Gamma and Tau_2-(CT187)
           305233
                      304340
                                 R
CPn0273
           305844
                      305227
                                       tdk-Thymidylate Kinase-(CT188)
                                 R
                                       gyrA_1-DNA Gyrase Subunit A_1-(CT189)
CPn0274
           308353
                      305852
                                 R
CPn0275
                      308372
                                       gyrB_1-DNA Gyrase Subunit B_1-(CT190)
           310786
                                 R
CPn0276
           311137
                      310793
                                       CT191 hypothetical protein
                                 R
CPn0277
           311910
                      311404
                                 R
                      312060
CPn0278
           312875
                                       *conserved outer membrane lipoprotein protein
                                 R
CPn0279
           313537
                      312875
                                       *Possible ABC Transporter Permease Protein
CPn0280
           314572
                      313550
                                       dppF-Dipeptide Transporter ATPase-(CT689)
                                 R
```

```
dhnA-Predicted 1,6-Fructose Biphosphale Aldolase (dehydrin family) -
CPn0281
         315057
                     316103
                                 F
                                         (CT215)
                                      xasA/gadC-Amino Acid Transporter-(CT216)
                     317529
CPn0282
          316126
                                 F
         318497
                     317532
                                 R
CPn0283
CPn0284
          319045
                     318551
                     319051
          320595
                                 R
CPn0285
                                      mgtE-Mg++ Transporter (CBS Domain) - (CT194)
CPn0286
          322059
                      320650
                                 R
          324221
                     322089
                                 R
CPn0287
                                      CT195 hypothetical protein
                     324571
CPn0288
          325716
                                 R
                                      aaaT-Neutral Amino Acid (Glutamate) Transporter-(CT230)
          325812
                     326996
CPn0289
                                      Na-dependent Transporter-(CT231)
          327042
                     328523
                                F
CPn0290
                                      incB-Inclusion Membrane Protein B-(CT232)
                     329194
                                F
CPn0291
          328667
                                      incC-Inclusion Membrane Protein C-(CT233)
CPn0292
                      329836
                                 F
          329228
                                      CT234 hypothetical protein
                     332723
                                F
CPn0293
          329949
                                      CAMP-Dependent Protein Kinase Regulatory Subunit-(CT235)
                     333502
                                 F'
          333092
CPn0294
                                      acpP-Acyl Carrier Protein-(CT236)
CPn0295
          333863
                     333627
                                R
                                      fabG-Oxoacyl (Carrier Protein) Reductase-(CT237)
          334765
                     334022
                                 R
CPn0296
                                      fabD-Malonyl Acyl Carrier Transcyclase-(CT238)
                     334774
CPn0297
          335697
                                 :3
                                      fabH-Oxoacyl Carrier Protein Synthase III-(CT239)
          336721
                     335717
                                15
CPn0298
                                      recR-Recombination Protein-(CT240)
CPn0299
          336816
                     337415
                                17
                     340152
                                      yaeT-Omp85 Analog-(CT241)
                                F
          337783
CPn0300
                                      (OmpH-Like Outer Membrane Protein) - (CT242)
                      340762
CPn0301
          340250
                                Γ
                                      lpxD-UDP Glucosamine N-Acyltransferase-(CT243)
CPn0302
          340787
                     341866
                                Ŧ.
                                      CT244 hypothetical protein
          342958
                     341921
                                F!
©Pn0303
                                      pdhA/odpA-Pyruvate Dehydrogenase Alpha-(CT245)
          343133
                     344158
                                F
CPn0304
                                      pdhB/odpB-Pyruvate Dehydrogenase Beta-(CT246)
ÇPn0305
          344154
                     345137
                                I
                                      pdhC-Dihydrolipoamide Acetyltransferase-(CT247)
                      346431
CPn0306
          345145
                                7
                                      glgP-Glycogen Phosphorylase-(CT248)
                     346515
CPn0307
          348986
                                Ι.
                                      similarity to CT249
          349234
                     349596
@Pn0308
                                      dnaA_1-Replication Initiation Protein_1-(CT250)
⊈₽n0309
          350974
                     349595
                                R
                                      60IM-60kDa Inner Membrane Protein-(CT251)
ĈPn0310 · 353433
                     351049
                                R
                                      lgt-Prolipoprotein Diacylglycerol Transferase-(CT252)
dPh0311
          354438
                     353575
                                 R
                                      CT101 hypothetical protein
          354524
                     354976
                                F
@Pn0312
                                      acps-Acyl-carrier Protein Synthase-(CT100)
                     355355
CPn0313
          354990
                                 F
                                      trxB-Thioredoxin Reductase-(CT099)
          356285
                     355353
                                R
@Pn0314
                                      rs1-S1 Ribosomal Protein-(CT098)
ČPn0315
          356977
                      358716
                                F
                                      nusA-N Utilization Protein A-(CT097)
                     360121
CPn0316
          358820
                                F
                               F`
@₽n0317
          360081
                     362750
                                      infB-Initiation Factor-2-(CT096)
                                      rbfA-Ribosome Binding Factor A-(CT095)
CPn0318
          362767
                     363126
                                F
                                      truB-tRNA Pseudouridine Synthase-(CT094)
CEn0319
          363175
                     363879
                                F
                                      ribF-FAD Synthase-(CT093)
CPn0320
          363860
                     364783
                                F
                                      ychF-GTP Binding Protein-(CT092)
          365858
                     364767
CPn0321
                                R
                                      yscU-YopS Translocation Protein U -(CT091)
          366249
                     367328
CPn0322
                                      lcrD- Low Calcium Response D-(CT090)
CPn0323
          367331
                     369460
                                F
                                      lcrE- Low Calcium Response E-(CT089)
CPn0324
          369492
                      370688
                                F
                                      sycE-Secretion Chaperone-(CT088)
CPn0325
          370708
                     371148
                                F
                                      malQ-Glucanotransferase-(CT087)
CPn0326
          371148
                     372725
                                F
                                      rl28-L28 Ribosomal Protein-(CT086)
CPn0327
          372945
                     373211
                               F
CPn0328
          373241
                     374992
                                      CT085 hypothetical protein
                                F
                                      Phopholipase D Superfamily [leader (33) peptide]-(CT084)
CPn0329
           375088
                      376146
                                 F
                                      CT083 hypothetical protein
CPn0330
          376675
                     376202
                                R
          378437
                                      CT082 hypothetical protein
CPn0331
                     376701 · R
                                      CHLTR T2 Protein-(CT081)
          378655
CPn0332
                      378536
                                 R
           379090
CPn0333
                      378800
                                 R
                                      1tuB-(CT080)
CPn0334
           379311
                      379823
                                 F
                                      CT079 similarity
                                      folD-Methylene Tetrahydrofolate Dehydrogenase-(CT078)
CPn0335
          379817
                      380674
                                 F
CPn0336
          380650
                      381591
                                 F
                                      yojL-(CT077)
CPn0337
           382027
                                      smpB- Small Protein B-(CT076)
                     381575
                                 R
                                      dnaN-DNA Pol III (beta chain)-(CT075)
CPn0338
           382278
                      383375
                                 F
CPn0339
          383420
                                      recF-ABC superfamily ATPase-(CT074)
                      384034
                                 F
                                      (frame-shift with 0339)
CPn0340
         383842
                      384156
                                 F
          384160
                                      (frame-shift with 0340)
CPn0341
                      384495
                                 F
                                      predicted OMP [leader (19) peptide]-(CT073)
CPn0342
          384622
                      385062
                                 F
                                      (frame-shift with 0342?)
CPn0343
           .84999
                      385595
                                 F
                                      yaeL-Metalloprotease-(CT072)
CPn0344
           387420
                      385558
                                 R
CPn0345
          388572
                      387436
                                      yaeM-(CT071)
                                 R
                                      troD/ytgD-Integral Membrane Protein-(CT070)
CPn0346
          389675
                      388704
                                      troC/ytgC-Integral Membrane Protein-(CT069)
CPn0347
           391021
                      389678
                                 R
                                      troB/ytgB-ABC transporter ATPase-(CT068)
CPn0348
                      391027
           391803
                                 R
                                      troA/ytgA-Solute Protein Binding Family-(CT067)
CPn0349
          392770
                      391790
                                 R
CPn0350
           393181
                      393684
                                      CT066 hypothetical protein
                                      adt_1-ADP/ATP Translocase_1-(CT065)
                                 F
CPn0351
           393888
                      395432
```

```
CPn0352
           395574
                       396830
                       397135
CPn0353
           396893
CPn0354
           397167
                       398507
                                  F
           399889
                       398591
                                  R
CPn0355
           400459
                       400109
CPn0356
           401317
                       400469
CPn0357 .
           401751
                       401578
                                  R
CPn0358
                                        lepA-GTPase-(CT064)
           402012
                       403817
                                  F
CPn0359
                                        gnd-6-Phosphogluconate Dehydrogenase-(CT063)
           405358
                       403922
                                  R
CPn0360
                                        tyrs-tyrosyl tRNA Synthetase-(CT062)
                       405382
           406647
                                  R
CPn0361
                                        fliA/rpsD-Sigma-28/WhiG Family-(CT061)
           407825
                       407055
CPn0362
                                       flhA-Flagellar Secretion Protein-(CT060)
CPn0363
           409688
                       407943
                                  R
                                        fer4-Ferredoxin IV-(CT059)
           409966
                       410238
                                  F
CPn0364
           410528
                       411544
                                  F
CPn0365
                       412440
                                  F
           411976
CPn0366
CPn0367
           413102
                       413836
                                  F
           413790
                                  F
                       414107
CPn0368
CPn0369
           414351
                       415562
                                  F
                                       CT058 hypothetical protein_2
                                       CT058 hypothetical protein_3
CPn0370
           415800
                       416912
                                  F
           417147
                       417503
                                  F
CPn0371
           417687
                       418001
C₽:20372
                                  F
CPn0373
                       420218
                                  F
                                       gcpE-(CT057)
           418380
                       420961
                                       CT056 hypothetical protein
CPn0374
           420218
                                  F
                       421615
                                  F
CPn0375
           421121
CPn0376
           421854
                       422294
                                  F
                                       sucB_1-Dihydrolipoamide Succinyltransferase_1-(CT055)
CPn0377
           423438
                       422347
                                  R
                                       sucA-Oxoglutarate Dehydrogenase-(CT054)
CPn0378
           426168
                      423445
                                  R
                                       CT053 hypothetical protein
CPn0379
           426322
                       426765
                                  F
                                       hemN_1-Coproporphyrinogen III Oxidase_1-(CT052)
           426758
                       427876
CPn0380
                                  F
           429809
                       428037
                                       CT326 similarity
CPn0381
                                       yabC/yraL-SAM-Dependent Methytransferase-(CT048)
           430749
CPn0382
                      430036
                                  R
                                       CT047 hypothetical protein
CPn0383
           431693
                       430749
                                  R
                                       hctB-Histone-like Protein 2-(CT046)
           432377
CP110384
                      431862
                                  R
                                       pepA-Leucyl Aminopeptidase A-(CT045)
CPn0385
           434018
                       432522
                                  R
                                       ssb-SS DNA Binding Protein-(CT044)
CPn0386
           434525
                      434046
                                  R
CPn0387
           435196
                       434699
                                       CT043 hypothetical protein
                                  R
                                       glgX-Glycogen Hydrolase (debranching)-(CT042)
CPn0388
           435329
                       437320
                                  F
           438134
                       437319
                                       CT041 hypothetical protein
                                  RÌ
CPm0389
CP00390
                                       ruvB-Holliday Junction Helicase-(CT040)
           439144
                       438134
           439692
                      439510
CPn0391
                                  R
                                       dcd-dCTP Deaminase-(CT039)
CPn0392
           439814
                       440383
                                  F
                                       CT038 hypothetical protein
           440379
                       440723
CPn0393
                                  F
                                       tlyC_1-CBS Domain protein (Hemolysin Homolog)_1-(CT256)
CPn0394
           440736
                       441968
                                  F
CPn0395
           441964
                       443175
                                  F
                                       CT257 hypothetical protein
                                       yhfo-NifS-related protein-(CT258)
CPn0396
           444353
                      443241
                                  R
                                       PP2C phosphatase family-(CT259)
CPn0397
           445115
                       444381
                                  R
CPn0398
           445533
                       445700
                                  F
CPn0399
           445879
                       446523
                                       CT253 hypothetical protein
CPn0400
           446536
                      447306
                                 F
                                       CT254 hypothetical protein
CPn0401
           447884
                       447495
                                  R
                                       CT255 hypothetical protein
                                       mutY-Adenine Glycosylase-(CT107)
           448994
CPn0402
                       447888
                                  R
                                       yceC-predicted pseudouridine synthetase family-(CT106)
CPn0403
           449015
                       449710
                                  F
CPn0404
           450887
                       449871
CPn0405
           451739
                       450966
                                       CT105 hypothetical protein
                                  R
                                       fabI-Enoyl-Acyl-Carrier Protein Reductase-(CT104)
CPn0406
           451969
                       452865
                                  F
                                       HAD superfamily hydrolase/phosphatase-(CT103)
CPn0407
           453742
                       452858
                                  R
CPn0408
           454105
                       454581
                                  F
                                       CT102 hypothetical protein
CPn0409
           454645
                       455127
                                       CT260 hypothetical protein
                                       dnaQ_1-DNA Pol III Epsilon Chain_1-(CT261)
CPn0410
           455123
                       455833
                                  F
CPn0411
           455833
                       456609
                                  F
                                       CT262 hypothetical protein
                                       CT263 hypothetical protein
CPn0412
           456590
                       457246
                                  F
                                       msbA-Transport ATP Binding Protein-(CT264)
CPn0413
           459203
                       457227
                                       accA-AcCoA Carboxylase/Transferase Alpha-(CT265)
CPn0414
           460143
                      459172
                                  R
CPn0415
           461498
                       460221
                                  R
                                       CT266 hypothetical protein
                                       himD/ihfA-Integration Host Factor Alpha-(CT267)
CPn0416
           461856
                       461557
                                  R
                                       amiA-N-Acetylmuramoyl Alanine Amidase-(CT268)
CPn0417
           463035
                       462244
                                  R
                                       murE-N-Acetylmuramoylalanylglutamyl DAP Ligase-(CT269)
CPn0418
           464401
                       462953
                                  R
CPn0419
                                       pbp3- transglycolase/transpeptidase-(CT270)
           466834
                       464876
                                  R
CPn0420
           467108
                                       CT271 hypothetical protein
                       466824
                                       yabC-PBP2B Family methyltransferase-(CT272)
                       467108
CPn0421
           467998
                                  R
CPn0422
           4.68242
                       468784
                                  F
                                       CT273 hypothetical protein
CPn0423
           468791
                       469216
                                  F
                                       CT274 hypothetical protein
```

```
dnaA_2-Replication Initiation Factor_2-(CT275)
CPn0424
           469612
                      470961
                                        CT276 hypothetical proteins
CPn0425
           470980
                       471564
                                  F
                       471536
                                  R
                                        CT277 similarity
           472111
CPn0426
                                        nqr2-NADH (Ubiquinone) Dehydrogenase-(CT278)
                       473715
                                  F
CPn0427
           472207
                                       nqr3-NADH (Ubiquinone) Oxidoreductase, Gamma-(CT279)
                      474681
                                  F
CPn0428
           473722
                                       nqr4-NADH (Ubiquinone) Reductase 4-(CT280)
           474681
                       475319
CPn0429
                                        ngr5-NADH (Ublquinone) Reductase 5-(CT281)
                                  F
           475326
                      476093
CPn0430
                      476151
                                  R
           476483
CPn0431
                       476514
                                  R
           476816
CPn0432
                                        gcsH-Glycine Cleavage System H Protein-(CT282)
                      476929
                                  R
CPn0433
           477273
                                        CT283 hypothetical protein
           479462
                      477276
CPn0434
                                        Phospholipase D superfamily [uncleavable leader peptide]-(CT284).
           480902
                      479475
                                  R
CPn0435
                                        lplA-Lipoate Protein Ligase-Like Protein-(CT285)
           481618
                      480902
                                  R
CPn0436
                                        clpC-ClpC Protease-(CT286)
                      484350
                                  F
_CPn0437
           481816
                                       ycbF-PP-loop superfamily ATPase-(CT287)
                      484334
                                  R
CPn0438
           485416
           485553
                       486077
                                  F
CPn0439
CPn0440
                      486740
                                  F
           486105
           486891
                      487838
                                  F
                                       CT007 hypothetical protein
CPn0441
                                       CT006 hypothetical protein
           488013
                       488528
                                  F
CPn0442
                      489979
                                  F
                                       CT005 hypothetical protein
           488729
CPn0443
                                       pmp_6-Polymorphic Outer Membrane Protein G/I Family
                      494507
CPn0444
           490287
                                        pmp_7-Polymorphic Outer Membrane Protein G Family
                      497579
                                  F
CPn:0445
           494772
                                        pmp_8-Polymorphic Outer Membrane Protein G Family
CPn0446
                       500415
                                  F
           497626
                                        pmp_9-Polymorphic Outer Membrane Protein G/I Family
CPn0447
           500568
                       503351
                                  F
                                        *yxjG_Bs_2 Hypothetical Protein
           504810
                      503698
                                  R
CPm0448
                                        pmp_10-PMP_10 (Frame-shift with 0451)
                       505330
CPn0449
           507231
                                  R
                                        pmp_10-Polymorphic Outer Membrane Protein G Family
                      507180
CP110450
           508112
                                  R
                                        pmp_11-Polymorphic Outer Membrane Protein G Family
CPn0451
                                  F
           508275
                       511058
                                        pmp_12-Polymorphic Outer Membrane Protein A/I Family (truncated)
CPn0452
           511319
                      512860
                                  F
                                       pmp_13 -Polymorphic Outer Membrane Protein G Family
           513234
                      516152
                                  F
CPn0453
                                        pmp_14-Polymorphic Outer Membrane Protein H Family
CPn0454
           516182
                      519115
                                  F
                      519458
                                  R
CPn0455
           520348
CPn0456
           521532
                      520327
                                  R
CPn0457
           523865
                      522120
                                  R
CPn 0458
           526320
                      524236
                                  R
CPn0459
           527005
                      526619
                                  R
           527840
                       526992
CPn0460
                                  R
           528638
                      527844
CPn0461
                                  R
                       529037
CPm0462
           531052
                                  R
CPn0463
           532357
                      531191
                                  R
CPn0464
           532842
                       532366
                                  R
CPn0465
           533212
                      532871
                                  R
                                        pmp_15-Polymorphic Outer Membrane Protein E Family
           533724
                      536537
CPn0466
                                  F
                                        pmp_16-Polymorphic Outer Membrane Protein E Family
CPn0467
           536633
                      539434
                                  F
                                        pmp_17-Polymorphic Outer Membrane Protein E Family
                       540432
CPn0468
           539632
                                  F
                                        pmp_17-Polymorphic Outer Membrane Protein (Frame-shift with 0469)
           540399
                       541460
                                  F
CPn0469
                                        pmp_17-Polymorphic Outer Membrane Protein (Frame-shift with 0470)
                                  F
CPn0470
           541357
                       542532
                                        pmp_18-Polymorphic Outer Membrane Protein E/F Family
CPn0471
           542564
                       545401
                                  F
CPn0472
           547905
                      545581
                                  R
CPn0473
           549593
                       548070
                                  R
                                        CT365 hypothetical protein
CPn0474
           551573
                       549807
                                  R
                                        glgB-Glucan Branching Enzyme-(CT866)
CPn0475
                       551685
           553844
                                  R
CPn0476
           554844
                       553858
                                        CT865 hypothetical protein
                                        *yqeV_Bs Hypothetical Protein
CPn0477
           556106
                       554844
                                  R
                                        hflX-GTP Binding Protein-(CT379)
CPn0478
           557625
                       556210
                                  R
                                        phnP-Metal Dependent Hydrolase-(CT380)
CPn0479
           558425
                       557616
                                  R
                                        CT383 hypothetical protein
CPn0480
           559303
                       558650
                                  R
CPn0481
           560946
                       559339
                                  R
                                        artJ-Arginine Periplasmic Binding Protein-(CT381)
CPn0482
           561737
                       560961
                                  R
CPn0483
           561836
                       564964
                                  F
                                        aroG-Deoxyheptonate Aldolase-(CT382)
CPn0484
           564970
                       565824
                                  F
                                        CT382.1 hypothetical protein
CPn0485
           566038
                       566229
                                  F
                                        *hypothetical proline permease
           567784
                       566405
CPn0486
                                  R
                                        CT384 hypothetical protein
CPn0487
           569740
                       568112
                                  R
                                        hitA-HIT Family Hydrolase-(CT385)
CPn0488
           570096
                       569767
                                  R
                                        CT386 hypothetical protein
CPn0489
           570965
                       570096
                                  R
CPn0490
                                        CT387 hypothetical protein
            571279
                       573333
                                        CT389 hypothetical protein
CPn0491
                                  R
           574352
                       573336
CPn0492
           574652
                       574804
                                  F
CPn0493
            575004
                       574855
                                  R
                       575146
CPn0494
           575364
                                  R
                                        aspC-Aspartate Aminotransferase-(CT390)
           575603
                       576793
CPn0495
```

```
577812
                                       CT391 hypothetical protein
CPn0496
           576793
                                       CT388 hypothetical protein
CPn0497
           578089
                      577820
                      578085
           579035
                                 R
CPn0498
CPn0499
           580359
                      579205
                                       proS-Prolyl tRNA Synthetase-(CT393)
CPn0500
           580659
                      582362
                                 F
                                       hrcA-HTH Transcriptional Repressor-(CT394)
CPn0501
           582457
                      583650
                                       grpE-HSP-70 Cofactor-(CT395)
                      584201
                                 F
CPn0502
           583650
                                 F
                                       dnaK-HSP-70-(CT396)
          584234
                      586213
CPn0503
                                       vacB-ribonuclease family-(CT397)
           586487
                      588514
                                 F
CPn0504
                                       *3-methyladenine DNA glycosylase
CPn0505
           588519
                      589106
                                 F
                                       CT421 hypothetical protein
                      589840
                                 F
CPn0506
           589172
                                       CT421.1 hypothetical protein
                      590122
                                 F
CPn0507
           589961
                      590300
                                 F
                                       CT421.2 hypothetical protein
CPn0508
           590142
                                       (predicted Metalloenzyme) - (CT422)
                      590808
                                 F
CPn0509
           590335
                                       tlyC_2-CBS Domains (Hemolysin homolog)_2-(CT423)
                                 F
                      591973
CPn0510
           590813
           592141
                      592488
                                       rsbV_1-Sigma Regulatory Factor_1-(CT424)
CPn0511
                                       CT425 hypothetical protein
                                 F
CPn0512
           592553
                      594412
                      595753
                                 F
                                       Fe-S oxidoreductase_1-(CT426)
           594647
CPn0513
                                       CT427 hypothetical protein
           595729
                      596520
                                 F
CPn0514
                                       ubiE-Ubiquinone Methyltransferase-(CT428)
                      597181
                                 F
           596492
CPn0515
                      597255
CPn0516
           598814
                                 R
CPn0517
           599631
                      598795
                                 R
                                       CT429 hypothetical protein
           600803
                      599832
                                 R
CPn0518
                                       dapF-Diaminopimelate Epimerase-(CT430)
                      600904 .
CPn0519
           601674
                                 R
                                       clpP-CLP Protease-(CT431)
                      601646
           602218
                                 R
CPn0520
                                       glyA-Serine Hydroxymethyltransferase-(CT432)
                      602241
CPn0521
           603797
                                       CT433 hypothetical protein
                                 F
CPn0522
           603987
                      604655
                                 F
CPn0523
           604723
                      605052
CPn0524
           605103
                      606179
                                 F
                                       CT398 hypothetical protein
                      607283
           606522
CPn0525
                                       yrbH-GutQ/KpsF Family Sugar-P Isomerase-(CT399)
CPn0526
           608696
                      607710
                                       sucB_2-Dihydrolipoamide Succinyltransferase_2-(CT400)
C₽±0527
           609904
                      608726
                                 R
                                       gltT-Glutamate Symport-(CT401)
CPn0528
           611162
                      609921
                                 R
                                       ycaH-ATPase-(CT402)
CPn0529
           612259
                      611165
                                 R
CPn0530
           613254
                      612460
                                       spou_1-rRNA Methylase_1-(CT403)
                                       SAM dependent methyltransferase-(CT404)
                      613245
CPn0531
           614069
                                 R
                                       ribC/risA-Riboflavin Synthase-(CT405)
CPn0532
           614674
                      614075
                                 R
                                       CT406 hypothetical protein
CPn0533
           614930
                      615385
                                 F
                                       dksA-DnaK Suppressor-(CT407)
CPn0534
           615413
                      615784
                                 F
                                       lspA-Lipoprotein Signal Peptidase-(CT408)
CPn0535
           615793
                      616296
                                 F
                                       dagA_1-D-Ala/Gly Permease_1-(CT409)
CPn0536
           616345
                      617691
                                 F
                                       CT814.1 hypothetical protein
CPn0537
           617833
                      618189
                                 F
                                       CT814 hypothetical protein
                      618511
CPn0538
           618212
                                       pmp_19-polymorphic outer membrane protein A Family -(CT412)
CPn0539
           618705
                      621545
                                       pmp_20-polymorphic outer membrane protein B Family-(CT413)
CPn0540
           621694
                      626862
                                 F
                                       Solute binding protein (-yebL-Synechocystis Adhesin Homolog)-(CT415)
           627170
                      628003
                                 F
CPn0541
                                       ABC Transporter ATPase-(CT416)
CPn0542
           628003
                      628737
                                 F
                                      (Metal Transport Protein) - (CT417)
CPn0543
           628725
                      629603
                                       yhbZ-GTP binding protein-(CT418)
CPn0544
           630529
                      629525
                                 R
                                       rl27-L27 ribosomal protein-(CT419)
CPn0545
           630884
                      630633
                                 R
                                       rl21-L21 Ribosomal Protein-(CT420)
CPn0546
           631229
                      630912
                                 R
                                       ygbB family-(CT434)
CPn0547
           631661
                      632188
                                F
                                       cysJ-Sulfite Reductase-(CT435)
CPn0548
           633231
                      632191
                                 R
                    • 633255
                                       rs10-S10 Ribosomal Protein-(CT436)
CPn0549
           633569
                                 R
                                       fusA-Elongation Factor G-(CT437)
CPn0550
           635661
                      633580
                                 R
                                       rs7-S7 Ribosomal Protein-(CT438)
CPn0551
           636168
                      635698
                                 R
                                       rs12-S12 Ribosomal Protein-(CT439)
CPn0552
           636587
                      636219
                                 R
CPn0553
           637747
                      636812
                                       CT440 hypothetical protein
CPn0554
           637854
                      638141
                                 F
                                       tsp-Tail-Specific Protease-(CT441)
CPn0555
           638298
                      640241
                                 F
                                       crpA-15kDa Cysteine-Rich Protein-(CT442)
CPn0556
           640912
                      640325
                                 R
                                       omcB-60kDa Cysteine-Rich Outer Membrane Complex Protein-(CT443)
CPn0557
           642861
                      641194
                                 R
                                       omcA-9kDa-Cysteine-Rich Outer Membrane Complex Lipoprotein-(CT444)
CPn0558
           643300
                      643031
                                 R
CPn0559
                      643927
                                 F
                                       CT441.1 hypothetical protein
           643742
                                       gltX-Glutamy1-tRNA Synthetase-(CT445)
CPn0560
           645612
                      644098
                                 R
                                       euo-CHLPS Euo Protein-(CT446)
CPn0561
                      645871
           646404
                                 R
                                       *CHLPS 43 kDa protein homolog_1
CPn0562
           648036
                      646918
                                 R
                                       recJ-ssDNA Exonuclease-(CT447)
CPn0563
                      648293
           650056
                                 R
                                       secD&secF-Protein Export Proteins SecD/SecF (fusion)-(CT448)
CPn0564
           654350
                      650145
                                 R
CPn0565
           655630
                      654533
                                 R
                                       CT449 hypothetical protein
CPn0566
                      656890
                                       yaeS family-(CT450)
           656141
                                 F
                                       cdsA-Phosphatidate Cytidylytransferase-(CT451)
CPn0567
           656894
                      657817
```

```
cdsA-Phosphatidate Cytidylytransferase-(CT452)
CPn0568
           657817
                       658464
                                       plsC-Glycerol-3-P Acyltransferase-(CT453)
           658464
                      659099
CPn0569
                                       args-Arginyl tRNA Transferase-(CT454)
CPn0570
           659107
                       660789
                                  F
                                       murA-UDP-N-Acetylglucosamine Transferase-(CT455)
           662122
                      660749
                                  R
CPn0571
                                       CT456 hypothetical protein
                       664616
           662352
CPn0572
                                       yebC family-(CT457)
           665404
                      664691
                                  R
CPn0573
           665945
                      665394
CPn0574
                                       Yhhy-Amino Group Acetyl Transferase-(CT458)
                       665982
                                  R
CPn0575
           666494
                                       prfB-Peptide Chain Release Factor 2 (natural UGA frame-shift )-(CT45
                       666494
                                  R
CPn0576
           667543
                                       prfB-(natural UGA frame-shift )
           667598
                       667530
                                  R
CPn0576
                                       SWIB (YM74) complex protein-(CT460)
CPn0577
           667895
                       668155
                                  F
                                       yaeI-phosphohydrolase-(CT461)
                       669365
                                  F
           668406
CPn0578
                                       ygbP/yacM-Sugar Nucleotide Phosphorylase-(CT462) .
CPn0579
           669361
                       669993
                                  F
                                        truA-Pseudouridylate Synthase I-(CT463)
                                  F
                       670793
CPn0580
           669993
                                       Phosphoglycolate Phosphatase-(CT464)
           671434
                       670745
                                  R
CPn0581
                                       CT465 hypothetical protein
                       672177
                                  F.
           671503
CPn0582
                                       CT466 hypothetical protein
           672400
                       672717
CPn0583
                                       atoS/ntrB-2-Component Sensor-(CT467)
CPn0584
           672707
                       673798
                                  ŀ,
                                        *similarity to Cps IncA_2
                       673865
                                  F.
CPn0585
           675817
                                        atoC/ntrC-2-Component Regulator-(CT468)
                       677183
           676026
                                  F.
CPn0586
                                        *yvyD_Bs conserved hypothetical protein
                       678124
                                  F
CPn0587
           677441
                                        CT469 hypothetical protein
           678084
                       678626
CPn0588
                                       CT470 hypothetical protein
           678640
                       679395
                                  F
CPa0589
                                       CT471 hypothetical protein
                       679516
           680112
                                  F
C₽ 10590
                                       yagE family-(CT472)
CPn0591
           680373
                       681020
                                  F
                                       yidD family-(CT473)
                       681461
CP00592
           681153
                                  F
                                        CT474 hypothetical protein
                       681391
C 200593
           682476
                                        pheT-phenylalanyl tRNA Synthetase Beta-(CT475)
                       684958
CPn0594
           682583
                                  F
                                        CT476 hypothetical protein
CPn0595
           684958
                       685926
                                  F
                                        ada-methyltransferase-(CT477)
CPn0596
           685939
                       686457
                                  F
                                        oppC_2-Oligopeptide Permease_2-(CT478)
                       686479
           688215
                                  R
CPn0597
                                        oppB_2-Oligopeptide Permease_2-(CT479)
CPn0598
           689697
                       688219
                                  R
                                        oppA_5-oligopeptide Binding Lipoprotein_5-(CT480)
                       689682
           691802
                                  R
GPn0599
                       691827
CPn0600
           692147
                                  R
                                        CT483 hypothetical protein
           693053
                       692736
CPn0601
                                  R
                                        CT484 hypothetical protein
                       693104
CPn0602
           694105
                                        hemZ-Ferrochetalase-(CT485)
                       695185
                                  F
CPn0603
           694205
                                        fliy-Glutamine Binding Protein-(CT486)
CPn0604
           695945
                       695196
                                  R
                                        yhhF-Methylase -(CT487)
CPn0605
           696707
                       696150
                                  R
                                        CT488 hypothetical protein
CPn0606
           697444
                       696707
                                  R
                                        glgC-Glucose-1-P Adenyltransferase-(CT489)
CPh0607
           698895
                       697573
                                  R
                                        *pyrF-Uridine 5'-Monophosphate Synthase (Ump Synthase)-truncated?
                       699016
CPn0608
           699645
                                  R
                                        CT490 hypothetical protein
           699705
                       699986
CPn0609
                                  F
                                        rho-Transcription Termination Factor-(CT491)
                       700029
CPn0610
           701420
                                  R
                                        yacE-predicted phosphatase/kinase-(CT492)
CPn0611
           702025
                       701420
                                        polA-DNA Polymerase I-(CT493)
CPn0612
           704631
                       702022
                                  R
                                        sohB-Protease-(CT494)
           705656
                       704658
                                  R
CPn0613
                                        adt_2-ADP/ATP Translocase_2-(CT495)
CPn0614
           707402
                       705783
                                  R
                                        pgsA_1-Glycerol-3-P Phosphatidyltransferase_1-(CT496)
                       707634
CPn0615
           708137
                                  R
                                        dnaB-Replicative DNA Helicase-(CT497)
CPn0616
           708791
                       710137
                                  F
                                        gidA-FAD-dependent oxidoreductase-(CT498)
           710484
                       712316
                                  F
CPn0617
                                        1plA-Lipoate-Protein Ligase A-(CT499)
CPn0618
           712306
                       713010
                                  F
                                        ndk-Nucleoside-2-P Kinase-(CT500)
           713444
                       713013
CPn0619
                                  R
                                        ruvA-Holliday Junction Helicase-(CT501)
CPn0620
           714139
                       713519
                                  R
                                        ruvC-Crossover Junction Endonuclease-(CT502)
           714647
CPn0621
                       714144
                                  R
                                        CT503 hypothetical protein
CPn0622
           715752
                       714793
                                  R
                                        CT504 hypothetical protein
CPn0623
           716993
                       716163
                                  R
                                        gapA-Glyceraldehyde-3-P Dehyrogenase-(CT505)
           718015
                       717011
CPn0624
                                  R
                                        rl17-L17 Ribosomal Protein-(CT506)
CPn0625
           718485
                       718060
                                        rpoA-RNA Polymerase Alpha-(CT507)
CPn0626
           719616
                       718495
                                  R
                                        rsl1-S11 Ribosomal Protein-(CT508)
CPn0627
            720038
                       719640
                                  R
                                        rs13-S13 Ribosomal Protein-(CT509)
CPn0628
            720428
                       720063
                                  R
                                        secY-Translocase-(CT510)
           721857
                       720487
CPn0629
                                  R
                                        rl15-L15 Ribosomal Protein-(CT511)
                       721885
CPn0630
            722316
                                        rs5-S5 Ribosomal Protein-(CT512)
            722806
                       722312
CPn0631
                                  R
                                        rl18-L18 Ribosomal Protein-(CT513)
CPn0632
            723195
                       722827
                                  R
                                        rl6-L6 Ribosomal Protein-(CT514)
CPn0633
            723757
                       723209
                                  R
                                        rs8-S8 Ribosomal Protein-(CT515)
CPn0634
            724185
                       723787
                                  R
                                        rl5-L5 Ribosomal Protein-(CT516)
            724745
                       724206
CPn0635
                                        rl24-L24 Ribosomal Protein-(CT517)
            725082
                       724750
CPn0636
                                  R
                                        rl14-L14 Ribosomal Protein-(CT518)
                       725099
CPn0637
            725464
                                   Ŕ
                                        rs17-S17 Ribosomal Protein-(CT519)
                       725490
            725747
CPn0638
```

```
rl29-L29 Ribosomal Procein-(CT520)
CPn0639
           725958
                       725743
                                  R
                                       rl16-L16 Ribosomal Protein-(CT521)
                       725964
           726377
CPn0640
                                       rs3-S3 Ribosomal Protein-(CT522)
           727077
                       726409
CPn0641
                                       rl22-L22 Ribosomal Protein-(CT523)
           727428
                      727096
CPn0642
                                       rs19-S19 Ribosomal Protein-(CT524)
           727713
                       727450
                                  R
CPn0643
                                       rl2-L2 Ribosomal Protein-(CT525)
           728573
                       727722
                                  R
CPn0644
                                       rl23-L23 Ribosomal Protein-(CT526)
           728930
                       728598
                                  R
CPn0645
                                       rl4-L4 Ribosomal Protein-(CT527)
                      728950
           729621
CPn0646
                                  R
                                        rl3-L3 Ribosomal Protein-(CT528)
           730331
                      729657
CPn0647
                                       CT529 hypothetical protein
CPn0648
           731603
                      730605
                                  R
                                        fmt-Methionyl tRNA Formyltransferase-(CT530)
                      731710
                                  R
           732672
CPn0649
                                        lpxA-Acyl-Carrier UDP-GlcNAc -(CT531)
CPn0650
           733501
                       732665
                                  R
                                        fabZ-Myristoyl-Acyl Carrier Dehydratase-(CT532)
                       733517
                                  R
           733975
CPn0651
                                        lpxC-Myristoyl GlcNac Deacetylase-(CT533)
           734835
                       733990
CPn0652
                                        cutE-Apolipoprotein N-Acetyltransferase-(CT534)
           736490
                      734868
                                  R
CPn0653
                                       vdlD/yciA-acyl-CoA Thioesterase-(CT535)
           736967
                       736503
                                  R
CPn0654
                                       dnaQ_2-DNA Pol III Epsilon Chain_2-(CT536)
           737847
                       737101
                                  R
CPn0655
                      738048
                                  F
           737872
CPn0656
                                       yjeE (ATPase or Kinase)-(CT537)
           738473
                      738051
                                  R
CPn0657
                                       CT538 hypothetical protein
           739168
                       738455
                                  R
CPn0658
                       739838
                                  F
                                        trxA-Thioredoxin-(CT539)
CPn0659
           739533
                                       spoU_2-rRNA Methylase_2-(CT540)
CPn0660
           740327
                       739860
                                  R
                                       mip-FKBP-type peptidyl-prolyl cis-trans isomerase-(CT541)
           741100
                       740327
                                  R
CPM0561
                                        aspS-Aspartyl tRNA Synthetase-(CT542)
           742923
                      741172
                                  R
CPn0662
           744190
                      742901
                                  R
                                       hisS-Histidyl tRNA Synthetase-(CT543)
CPn0663
           744757
                       744557
                                  R
CPm0664
                                       uhpC-Hexosphosphate Transport -(CT544)
                      746365
           745001
                                  F
CPn0665
           746388
                      750107
                                        dnaE-DNA Pol III Alpha-(CT545)
CPn0666
                                       predicted OMP [leader (17)-(CT546)
CPm0667
           751058
                      750177
                                  R
                                       CT547 hypothetical protein
CPm0668
           751209
                       752162
                                  F
                                       CT548 hypothetical protein
CPn0669
           752179
                       752775
                                  F
                                       rsbW-sigma regulatory factor-histidine kinase-(CT549)
           752765
                      753196
                                  F
CPn0670
                                       CT550 hypothetical protein
           753630
                      753205
CPm0671
                                       dacF(pbp5)-D-Ala-D-Ala Caroxypeptidase-(CT551)
CPn0672
           753741
                      755048
                                  F
           755287
                      755463
                                  F
                                       CT552 hypothetical protein
CPn0673
                                       fmu-RNA Methyltransferase-(CT553)
                      755577
C₽#0674
           756668
                                  R
           757919
                      756768
                                       CT696 hypothetical protein
CP#0675
                                  R
                                       homologous to CT695
           759217
                       758051
                                  R'
CP:0676
CPn0677
           760401
                       759256
                                  R
                      760682
CPri0678
           761320
                                  R
                                       pgk-Phosphoglycerate Kinase-(CT693)
CPn0679
           762930
                       761725
                                  R
                                       ygo4-Phosphate Permease-(CT692)
CPn0680
           764248
                      762971
                                  R
           764929
                      764258
                                       CT691 hypothetical protein
CPn0681
                                  R
                                        dppD-ABC ATPase Dipeptide Transport-(CT690)
           764984
                       765955
                                  F
CPn0682
                                       dppF-ABC ATPase Dipeptide Transport-(CT689)
CPn0683
                      766919
           765948
                                  F
                                        spoJ/parB-Chromosome Partitioning Protein-(CT688)
CPn0684
           768038
                       767181
                                  R
           768068
                      768217
CPn0685
                                  F
CPn0686
           768361
                       768176
                                  R
                                       CT482 hypothetical protein
CPn0687
           768564
                       769214
                                  F
CPn0688
           769382
                       770137
                                       CT481 hypothetical protein
                                  F
                                       yfhO_1-NifS-related Aminotransferase_1-(CT687)
CPn0689
           771404
                      770187
                                       ABC Transporter Membrane Protein-(CT686)
CPn0690
           772680
                       771436
                                  R
CPn0691
           773452
                       772685
                                  R
                                        abcX-ABC Transporter ATPase-(CT685)
CPn0692
           774912
                      773461
                                       ABC Transporter-(CT684)
                                  R
                                        TPR Repeats (O-Linked GlcNAc Transferase similarity)-(CT683)
CPn0693
           776256
                       775240
                                  R
                                       pbp2-PBP2-transglycolase/transpeptidase-(CT682)
CPn0694
           779599
                       776330
                                  R
                                        ompA-Major Outer Membrane Protein-(CT681)
CPn0695
           780216
                       781382
                                  F
                                        rs2-S2 Ribosomal Protein-(CT680)
CPn0696
           781769
                       782599
                                  F
                                        tsf-Elongation Factor TS-(CT679)
CPn0697
           782602
                       783447
                                  F
CPn0698
           783458
                       784201
                                  F
                                       pyrH-UMP Kinase-(CT679)
                                        rrf-Ribosome Releasing Factor-(CT677)
CPn0699
           784182
                       784721
                                  F
                                        CT676 hypothetical protein
CPn0700
           785097
                       785609
                                  F
                                        karG-Arginine Kinase-(CT675)
CPn0701
           785599
                       786672
                                  F
                                       YSCC/qspD-Yop C/Gen Secretion Protein D-(CT674)
CPn0702
           789685
                       786929
                                  R
                                        pkn5-S/T Protein Kinase-(CT673)
CPn0703
           791190
                       789685
                                        flin- Flagellar Motor Switch Domain/YscQ family-(CT672)
CPn0704
                       791209
           792321
                                  R
                                        CT671 hypothetical protein
CPn0705
           793173
                       792334
                                  R
                                        CT670 hypothetical protein
CPn0706
           793683
                       793180
                                  R
                                       yscN-Yop N (Flagellar-Type ATPase) - (CT669)
CPn0707
           795029
                       793704
                                  R
CPn0708
           795705
                       795034
                                        CT668 hypothetical protein
                                  R
CPn0709
           796188
                       795742
                                  R
                                        CT667 hypothetical protein
                                        CT666 hypothetical protein
CPn0710
           796461
                       796210
```

```
CPn0711
           796731
                       796486
                                        CT665 hypothetical protein
                                  R
                                        FHA domain; homology to adenylate cyclase)-(CT664)
CPn0712
           799315
                       796781
           799721
                                        CT663 hypothetical protein
CPn0713
                       799332
                                  R
                                        hemA-Glutamyl tRNA Reductase+(CT662)
CPn0714
           801107
                       800091
                                  R
                                        gyrB_2-DNA Gyrase Subunit B_2-(CT661)
CPn0715
           801657
                       803462
                                  F
                                        gyrA_2-DNA Gyrase Subunit A_2-(CT660)
CPn0716
           803469
                       804902
                                        CT656 hypothetical protein
           805010
                       805306
CPn0717
                                        CT657 hypothetical protein
           805309
                       805626
                                  F
CPn0718
                                        sfhB-(Pseudouridine Synthase)-(CT658)
                       806890
           805916
                                  F
CPn0719
                                        CT659 hypothetical protein
CPn0720
           807003
                       807236
                                  F
                                        kdsA-KDO Synthetase-(CT655)
                       808489
           807683
CPn0721
                                        CT654 hypothetical protein
           808489
                       808974
                                  F
CPn0722
                                        yhbG-ABC Transporter ATPase-(CT653)
                       809703
                                  F
           808984
CPn0723
                       809706
CPn0724
           810527
                                  R
                                        CT652.1 hypothetical protein
                       810587
CPn0725
           810811
                                  R
                                        CT620 hypothetical protein
           813372
                       810880
                                  R
CPn0726
                                        CT619 hypothetical protein
CPn0727
           813577
                       816192
                                  F
           818477
                       816525
                                  R
                                        CHLPN 76kDa Homolog_1 (CT622)
CPn0728
                                        CHLPN 76kDa Homolog_2 (CT623)
           819857
                       818592
CPn0729
                                  R
                       819963
                                        mviN-Integral Membrane Protein-(CT624)
           821603
CPn0730
                                  R
CPn0731
           821587
                       821760
                       822976
                                        nfo-Endonuclease IV-(CT625)
CPn0732
           822098
                                  F
                                        rs4-S4 Ribosomal Protein-(CT626)
CP110733
           823727
                       823101
                                  R
CP:0734
           823944
                       824915
                                  F
                                        vceA-(CT627)
                                        *pyrH/udk-Uridine Kinase (Uridine Monophosphokinase) (Pyrimidine
           825668
                       825003
CPn0735
                                  R
                                           Ribonucleoside Kinase).
                                        ygeD-Efflux Protein-(CT641)
CPn0736
                       825992
           827686
                                  R
                                        recC-Exodeoxyribonuclease V, Gamma-(CT640)
CP10737
           827685
                       830756
                                  F
                                        recB-Exodeoxyribonuclease V, Beta-(CT639)
CPm0738
           830746
                       833895
                                  F
CPn0739
           834871
                       833861
                                        CT638 hypothetical protein
                                  R
                                        tyrB-Aromatic AA Aminotransferase-(CT637)
CPn0740
           836048
                       834864
                                        greA-Transcription Elongation Factor-(CT636)
CPn0741
           838350
                       836185
                                  R
CPn0742
           838463
                       838888
                                  F
                                        CT635 hypothetical protein
                                        ngrA-Ubiquinone Oxidoreductase, Alpha-(CT634)
                       840362
CPn0743
           838962
                                  F
CPn0744
           841384
                       840389
                                        hemB-Porphobilinogen Synthase-(CT633)
           841903
                       841742
                                  R
CP110745
CPm0746
           841975
                       843567
                                  F
                                        CT632 hypothetical protein
                                  F'
CPh0747
           843675
                       843740
                                        CT631 hypothetical protein
                                        CT631 hypothetical protein (frame-shift)
CPn0747
           843725
                       843910
                                  F
                                        ispA-Geranyl Transtransferase-(CT628)
CPH0748
           844987
                       844121
                                  R
                                        glmU-UDP-GlcNAc Pyrophosphorylase-(CT629)
CPn 0749
           845629
                       845006
                                  R
                                        tctD/cpxR-HTH Transcriptional Regulatory Protein + Receiver Doman-
CPn0750
           846411
                       845707
                                  R
                                           (CT630)
CPn0751
           846608
                       848434
                                        CT651 hypothetical protein
                                  F
                                        recD_2-Exodeoxyribonuclease V, Alpha_2-(CT652)
CPn0752
           848604
                       850082
                                  F
CPn0753
           851006
                       850161
                                  R
                                        rs20-S20 Ribosomal Protein-(CT617)
CPn0754
           851336
                       851040
                                  R
CPn0755
           851597
                       852799
                                  F
                                        CT616 hypothetical protein
                                        rpoD-RNA Polymerase Sigma-66 -(CT615)
CPn0756
           852961
                      854676
                                  F
                                        folX-Dihydroneopterin Aldolase-(CT614)
CPn0757
           854733
                       855134
                                  F
CPn0758
           855110
                       856459
                                  F
                                        folP/dhpS-Dihydropteroate Synthase-(CT613)
CPn0759
           856488
                                        folA-Dihydrofolate Reductase-(CT612)
                       856997
                                  F
CPn0760
           856957
                       857694
                                  F
                                        CT611 hypothetical protein
CPn0761
           857704
                       858375
                                  F
                                        CT610 hypothetical protein
                                        recA-RecA recombination protein-(CT650)
CPn0762
           859597
                       858539
                                  R
                                        ygfA-Formyltetrahydrofolate Cycloligase-(CT649)
CPn0763
           860511
                       859972
                                  R
CPn0764
           861807
                       860524
                                        CT648 hypothetical protein
                                  R
CPn0765
           862382
                       861801
                                        CT647 hypothetical protein
CPn0766
           863782
                                        CT646 hypothetical protein
                       862394
                                  R
                                        CT645 hypothetical protein
CPn0767
           863884
                       864177
                                  F
                                        yohI/nir3-predicted oxidoreductase -(CT644)
CPn0768
           864159
                       865163
                                  F
                                        topA-DNA Topoisomerase I-Fused to SWI Domain-(CT643)
CPn0769
           867733
                       865121
                                  R
CPn0770
           868340
                       869131
                                  F
                                        CT642 hypothetical protein
CPn0771
                       869144
                                        rpoN-RNA Polymerase Sigma-54-(CT609)
           870463
                                  R
CPn0772
           872385
                       870469
                                        uvrD-DNA Helicase-(CT608)
                                  R
CPn0773
                                        ung-Uracil DNA Glycosylase-(CT607)
           872488
                       873195
                                  F
CPn0774
                                        CT606.1 hypothetical protein
           873195
                       873425
                                  F
CPn0775
                       873414
                                        yggV family-(CT606)
           874031
                                  R
CPn0776
           874246
                       875487
                                  F
                                        CT605 hypothetical protein
                                        groEL_2-heat shock protein-60 -(CT604)
CPn0777
           875601
                       877178
                                  F
CPn0778
           877505
                       878092
                                        tsa/ahpC-Thio-specific Antioxidant (TSA) Peroxidase-(CT603)
                                  F
CPn0779
           878481
                       878095
                                        CT602 hypothetical protein
```

```
papQ/amiB-N-Acetylmuramoyl-L-Ala Amidase-(CT601)
           879205
                      878591
CPn0780
                                       pal-Peptidoglycan-Associated Lipoprotein-(CT600)
                      879198
           879773
CPn0781
                                       tolB-polysaccharide transporter-(CT599)
           881065
                      879773
CPn0782
                                       CT598 hypothetical protein
                      881100
CPn0783
           881885
                                  R
                                       exbD-Biopolymer Transport Protein-(CT597)
           882296
                      881892
                                  R
CPn0784
                                       exbB/tolQ-polysaccharide transporter-(CT596)
           882991
                      882296
                                  R
CPn0785
                                       dsbD/xprA-Thio:disulfide Interchange Protein-(CT595)
CPn0786
           883185
                      885293
                                  F
                                       yabD/ycfH-PHP superfamily (urease/pyrimidinase) hydrolase-(CT594)
           885619
                      886401
CPn0787
                                       sdhC-Succinate Dehydrogenase-(CT593)
                                  F
           886542
                      887432
CPn0788
                                       sdhA-Succinate Dehydrogenase-(CT592)
           887439
                      889316
                                  F
CPn0789
                                       sdhB-Succinate Dehydrogenase-(CT591)
CPn0790
           889330
                      890103
                                  F
                                       CT590 hypothetical protein
                      890111
                                  R
CPn0791
           893050
                                       CT589 hypothetical protein
           894919
                      893108
                                  R
CPn0792
                                       rbsU-sigma regulatory family protein-PP2C phosphatase (RsbW
                                  Ŕ
           896823
                      894919
CPn0793
                                          antagonist) - (CT588)
           897174
                      898004
                                  F
CPn0794
                      899195 .
CPn0795
           898128
CPn0796
           899301
                      901340
                                  F
                                  F
                      902694
CPn0797
           901600
           902846
                      903856
                                  F
CPn0798
           904986
                      903940
                                  R
CPn0799
                                       eno-Enolase-(CT587)
                      905249
CPn0800
           906532
                                  R
                                       uvrB-Exinuclease ABC Subunit B-(CT586)
                      906727
CPn0801
           908697
                                  R
                                       trps-Tryptophanyl tRNA Synthetase-(CT585)
                      908709
CPn0802
           909740
                                  R
CPn0803
           910303
                      909752
                                  R
                                       CT584 hypothetical protein
                                       gp6D-CHLTR Plasmid Paralog-(CT583)
CPn0804
           911059
                      910310
                                  R
                                       minD-chromosome partitioning ATPase-CHLTR plasmid protein GP5D-(CT582)
           911831
                      911067
                                  R
C220805
                                       thrs-Threonyl tRNA Synthetase-(CT581)
CPn0806
           913771
                      911867
                                  R
                                       CT580 hypothetical protein
CPn0807
                      914879
                                  F
           913971
                                       CT579 hypothetical protein
CP::0808
           916287
                      914956
                                  R
                                       CT578 hypothetical protein
                      916307
           917785
                                  R
CPn0809
CPn0810
           918184
                      917825
                                       CT577 hypothetical protein
                                       lcrH_1-Low Ca Response Protein H_1-(CT576)
                      918208
                                  R
CPn0811
           918900
                                       mutL-DNA Mismatch Repair-(CT575)
CPn0812
           919123
                      920862
                                       pepP-Aminopeptidase P-(CT574)
CPn0813
           920870
                      921934
                                  F
                                       CT573 hypothetical protein
CPn0814
           922107
                      923357
                                  F
                                       gspD/pilQ-Gen. Secretion Protein D-(CT572)
CPm0815
           923361
                      925622
                                  F
                                       gspE-Gen. Secretion Protein E-(CT571)
           925615
                      927102
                                  F
CPn0816
                                       gspF-Gen. Secretion Protein F-(CT570)
CPn0817
           927115
                      928287
                                  F
                                       predicted OMP [leader (16) peptide]-(CT569)
                                  F
CPn0818
           928314
                      928682
                                       CT568 hypothetical protein
                      929132
CPn0819
           928689
                                       CT567 hypothetical protein
CPn0820
           929120
                      929659
                                  F
                                       CT566 hypothetical protein
CPn0821
           929667
                      930668
                                  F
                                       CT565 hypothetical protein
CPn0822
           930756
                      931229
                                  F
                                       yscT/spaR-YopT Tranlocation T-(CT564)
           932367
                      931501
                                  R
CPn0823
                                       yscS/fliQ-YopS/fliQ Translocation Protein-(CT563)
CPn0824
           932662
                      932378
                                  R
                                       yscR-Yop Translocation R-(CT562)
                      932677
CPn0825
           933594
                                  R
                                       yscL-Yop Translocation L-(CT561)
CPn0826
           934310
                      933612
                                  R
                                       CT560 hypothetical protein
           935264
                      934434
CPn0827
                                  R
                                       yscJ-Yop Translocation J-(CT559)
CPn0828
           936271
                      935267
CPn0829
           936744
                      937298
                                  F
CPn0830
           937444
                      937959
                                  F
CPn0831
           938267
                      938434
                                  F
                                       lipA-Lipoate Synthetase-(CT558)
CPn0832
           939747
                      938827
                                  R
                                       lpdA-Lipoamide Dehydrogenase-(CT557)
CPn0833
           941129
                      939747
CPn0834
           941553
                      942014
                                  F
                                       CT556 hypothetical protein
                                       motl_1-SWI/SNF family helicase_1-(CT555)
           945689
                      942045
CPn0835
                                       brnQ-Amino Acid (Branched) Transport-(CT554)
           946879
                      945722
CPn0836
                                  R
CPn0837
           947771
                      947145
                                  R
                                       nth-Enodnuclease III-(CT697)
                                       thdF-Thiophene/Furan Oxidation Protein-(CT698)
CPn0838
           949106
                      947781
                                  R
                                       psdD-Phosphatidylserine Decarboxylase-(CT699)
                      950159
                                  F
CPn0839
           949257
                      951544
                                       CT700 hypothetical protein
CPn0840
           950222
                                  F
                                       secA_2-Translocase SecA_2-(CT701)
                      954640
                                  F
CPn0841
           951731
                                       CT702 hypothetical protein (frame-shift with 0843)
CPn0842
           954883
                      954710
                                       CT702 hypothetical protein
CPn0843
           955191
                      954994
                                  R
CPn0844
                                       yphC-GTPase/GTP-binding protein-(CT703)
           956730
                      955270
                                  R
                                       pcnB_1-Poly A Polymerase_1-(CT704)
CPn0845
           958079
                       956850
           959374
                                       clpX-CLP Protease ATPase-(CT705)
CPn0846
                      958112
                                  R
                                       clpP-CLP Protease Subunit-(CT706)
CPn0847
           959995
                       959387
                                  R
                                       tig/murl-Trigger Factor-peptidyl-prolyl isomerase-(CT707)
CPn0848
           961502
                      960177
                                  R
                                       mot1_2-SWI/SNF family helicase_2-(CT708)
CPn0849
           961788
                       965285
                                  F
                                       mreB-Rod Shape Protein-Sugar Kinase-(CT709)
CPnC850
           965293
                       966390
```

```
pckA-Phosphoenolpyruvate Carboxykinase-(CT710)
CPn0851
           966396
                       968195
                                       CT711 hypothetical protein
CPn0852
           968316
                       970613
                                  F
CPn0853
           970637
                       971803
                                  F
                                       CT712 hypothetical protein
                                       ompB-Outer Membrane Protein B-(CT713)
           972837
CPn0854
                       971806
                                  Я
CPn0855
           973995
                       972994
                                       gpdA-Glycerol-3-P Dehydrogenase-(CT714)
           975377
                       973995
                                       AgX-1 Homolog-UDP-Glucose Pyrophosphorylase-(CT715)
CPn0856
           975757
                       975392
                                       CT716 hypothetical protein
CPn0857
                                  R
CPn0858
           977055
                       975757
                                  R
                                       fliI-Flagellum-specific ATP Synthase-(CT717)
CPn0859
           977588
                       977055
                                  R
                                       CT718 hypothetical protein
                       977608
                                       fliF-Flagellar M-Ring Protein-(CT719)
           978630
                                  R
CPn0860
           979722
                       978925
                                       nifu-Nifu-related protein-(CT720)
CPn0861
                       979722
                                       yfho_2-NifS-related protein_2-(CT721)
CPn0862
           980873
                                  Ŕ
                                       pgmA-Phosphoglycerate Mutase-(CT722)
CPn0863
           981514
                       980831
                                       yjbC-predicted pseudouridine synthase-(CT723)
CPn0864
           981670
                       982374
                                 F
CPn0865
           982418
                       982942
                                       CT724 hypothetical protein
CPn0866
           983491
                       982916
                                 R
                                       birA-Biotin Synthetase-(CT725)
           983423
                       984667
                                 F
                                       rodA-Rod Shape Protein-(CT726)
CPn0867
                                       zntA/cadA-Metal Transport P-type ATPase-(CT727)
CPn0868
           986643
                       984670
                                 \mathbf{F}^{j}
                                       CT728 hypothetical protein
           987401
                      986658
CPn0869
                                 F.
CPn0870
           988728
                       987448
                                       serS-Seryl tRNA Synthetase_2-(CT729)
           988772
                      989899
                                       ribD-Riboflavın Deaminase-(CT730)
CPn0871
                                 F
                                       ribA&ribB-GTP Cyclohydratase & DHBP Synthase -(CT731)
CPn0872
           989963
                       991216
                                 F,
@Pn0873
           991233
                      991694
                                       ribE-Ribityllumazine Synthase-(CT732)
                                       CT733 hypothetical protein
CPn0874
           993107
                      991749
CPh0875
           993372
                      994022
                                       CT734 hypothetical protein
                                       dagA_2-D-Alanine/Glycine Permease_2-(CT735)
@Pn0876
           994144
                      995517
                                 F
CPh0877
           995533
                       995982
                                 F
                                       ybcL family-(CT736)
           996654
                                       SET Domain protein-(CT737)
GPn0878
                      995992
                                 F
                                       yycJ-metal dependent hydrolase-(CT738)
GPn0879
           997439
                      996645
                                 R
CPn0880
           999861
                      997444
                                       ftsK-Cell Division Protein FtsK-(CT739)
           1005667
CPn0881
                      1006209
                                 F
CPh0882
           1006268
                      1007404
                                       dmpP/nqr6-Phenolhydrolase/NADH ubiquinone oxidoreductase-(CT740)
           1008865
CPn0883
                      1007573
                                 R
           1009359
CPn0884
                      1009009
                                 R
                                       CT741 hypothetical protein
CPn0885
           1010635
                      1009433
                                       ygcA-rRNA Methyltransferse-(CT742)
                                 R
CPh0886
           1011276
                      1010908
                                 R
                                       hctA-Histone-Like Developmental Protein-(CT743)
CPa0887
           1011692
                      1014157
                                 F
                                       CHLTR possible phosphoprotein-(CT744)
CPn0888
           1015423
                      1014119
                                 R`
                                       hemG-protoporphyrinogen Oxidase-(CT745)
CPn0889
           1016835
                                       hemN_2-Coproporphyrinogen III Oxidase_2-(CT746)
                      1015462
                                 R
CPn0890
           1017805
                      1016819
                                 R
                                       hemE-Uroporphyrinogen Decarboxylase-(CT747)
CPa0891
           1021073
                      1017819
                                       mfd-Transcription-Repair Coupling-(CT748)
                                 R
CPn0892
           1023661
                      1021046
                                       alaS-Alanyl tRNA Synthetase-(CT749)
                                 R
CPn0893
           1023894
                      1025888
                                       tktB-Transketolase-(CT750)
CPn0894
           1026766
                      1025888
                                 R
                                       amn-AMP Nucleosidase-(CT751)
CPn0895
           1026988
                      1027557
                                 F
                                       efp_2-Elongation Factor P_2-(CT752)
CPn0896
           1027595
                      1027822
                                 F
                                       CT753 hypothetical protein
CPn0897
           1028737
                      1027853
                                       (possible phosphohydrolase) - (CT754)
                                 R
CPn0898
           1030460
                      1028904
                                       Mitochondrial HSP60 Chaperonin Homolog-(CT755)
                                       murF-Muramoyl-DAP Ligase-(CT756)
CPn0899
           1030875
                      1032215
                                 F
CPn0900
           1032235
                      1033281
                                       mraY-Muramoyl-Pentapeptide Transferase-(CT757)
                                 F
CPn0901
           1033287
                                       murD-Muramoylalanıne-Glutamate Ligase-(CT758)
                      1034537
                                 F
CPn0902
           1034543
                      1035241 · F
                                       nlpD-Muramidase (invasin repeat family) - (CT759)
CPn0903
           1035263
                      1036417
                                 F
                                       ftsW-Cell Division Protein FtsW-(CT760)
CPn0904
           1036326
                      1037396
                                 F
                                       murg-Peptidoglycan Transferase-(CT761)
CPn0905
           1037409
                                       murC&ddlA-Muramate-Ala Ligase & D-Ala-D-Alam Ligase-(CT762)
                      1039835
                                 F
CPn0906
           1040340
                      1039915
                                       CT763 hypothetical protein
                                 R
                                       *CutA Periplasmic Divalent Cation Tolerance Protein CutA (C-Type
CPn0907
           1040780
                      1040445
                                          Cytochrome Biogenesis Protein)
CPn0908
           1041589
                      1040780
                                 R
                                       CT764 hypothetical protein
CPn0909
           1041637
                      1041966
                                       rsbV_2-Sigma Factor Regulator_2-(CT765)
                                 F
CPn0910
           1041979
                      1043004
                                       miaA-tRNA Pyrophosphate Transferase-(CT766)
CPn0911
           1044043
                      1042985
                                 R
                                       Fe-S cluster oxidoreductase_2-(CT767)
CPn0912
           1044129
                      1045760
                                 F
                                       CT768 hypothetical protein
CPn0913
           1045760
                      1045945
                                 F
CPn0914
           1045999
                      1046397
                                 F
CPn0915
                                       ybeB-iojap superfamily ortholog-(CT769)
           1046461
                      1046817
                                 F
CPn0916
           1046837
                      1048084
                                 F
                                       fabF-Acyl Carrier Protein Synthase-(CT770)
CPn0917
                                       hydrolase/phosphatase homolog-(CT771)
           1048090
                      1048539
                                 F
CPn0918
           1049223
                      1048579
                                       ppa-Inorganic Pyrophosphatase-(CT772)
                                 R
CPn0919
           1049378
                      1050430
                                       ldh-Leucine Dehydrogenase-(CT773)
                                 F
CPn0920
           1051405
                      1050431
                                       cysQ-Sulfite Synthesis/biphosphate phosphatase-(CT774)
CPn0921
           1051535
                      1052293
                                       snGlycerol-3-P Acyltransferase-(CT775)
```

```
aas-Acylglycerophosphoethanolamine Acydtransferase-(CT776)
CPn0922
            1052314
                       1053927
                                        bloF_1-Oxononanoate Synthase_1-(CT777)
CPn0923
            1053984
                       1055093
                                  F
                                        priA-Primosomal Protein N' -(CT778)
CPn0924
            1057274
                       1055028
                                  R
CPn0925
            1057900
                       1057226
                                  R
                                        CT779 hypothetical protein
           1058060
                       1058557
                                        Thioredoxin Disulfide Isomerase-(CT780)
CPn0926
            1059809
                                        *CHLPS 43 kDa protein homolog_2
CPn0927
                       1058670
           1061008
                       1059884
                                        *CHLPS 43 kDa protein homolog_3
CPn0928
                                  R
CPn0929
            1062292
                       1061186
                                  R
                                        *CHLPS 43 kDa protein homolog_4
CPn0930
            1062857
                       1063330
                                  F
                                        lysS-Lysyl tRNA Synthetase-(CT781)
                       1065718
           1064138
                                  F
CPn0931
                                        cysS-Cysteinyl tRNA Synthetase-(CT782)
CPn0932
           1067142
                       1065721
                                        predicted disulfide bond isomerase-(CT783)
CPn0933
           1067535
                       106857B
                                  F
                                        rnpA-Ribonuclease P Protein Component-(CT784)
CPn0934
           1068942
                       1068526
                                  R
                                        rl34-L34 Ribosomal Protein-(CT785)
CPn0935
           1069091
                       1068957
                                  R
                                        rl36-L36 Ribosomal Protein-(CT786)
                       1069470
CPn0936
           1069336
CPn0937
          .1069496
                       1069798
                                        rs14-S14 Ribosomal Protein-(CT787)
                                        CT788 hypothetical protein -[leader (60) peptide-periplasmic]
CPn0938
           1070322
                       1069849
                                  R
CPn0939
           1070728
                       1071195
                                  F
                                        CT790 hypothetical protein
                                        uvrC-Excinuclease ABC, Subunit C-(CT791)
CPn0940
           1073012
                       1071204
                                  R
CPn0941
           1075501
                       1073018
                                        mutS-DNA Mismatch Repair-(CT792)
                                  R
CPn0942
           1075985
                       1077754
                                        dnaG/priM-DNA Primase-(CT794)
           1077978
                       1078238
CPn0943
                                  F
                                        CT794.1 hypothetical protein
CPn0944
           1078512
                       1078997
                                  F
           1079070
CPm0945
                       1079660
                                  F
                                        CT795 hypothetical protein
CPn 0946
           1082786
                       1079745
                                        glyQ-Glycyl tRNA Synthetase-(CT796)
                                  R
CPn0947
           1083442
                       1084059
                                        pgsA_2-Glycerol-3-P-Phosphatydyltransferase_2-(CT797)
                                        glgA-Glycogen Synthase-(CT798)
CP#0948
           1085474
                       1084047
                                  R
CP10949
           1085929
                       1086483
                                        ctc-General Stress Protein-(CT799)
                                  F
                                        pth-Peptidyl tRNA Hydrolase-(CT800)
CPn0950
           1086488
                       1087027
                                  F
CPn0951
           1087122
                       1087457
                                       rs6-S6 Ribosomal Protein-(CT801)
                                  F
CPh@952
                                        rs18-S18 Ribosomal Protein-(CT802)
           1087478
                       1087723
                                  F
CPhi0953
           1087742
                       1088248
                                  F
                                       rl9-L9 Ribosomal Protein-(CT803)
           1088286
CPn0954
                       1088708
                                  F
                                       ychB-Predicted Kinase-(CT804)
CPn0955
           1088612
                       1089175
                                  F
                                        (frame-shift with 0954)
CPn0956
           1089560
                       1090909
                                  F
                                        CT805 hypothetical protein
CPn0957
           1093788
                       1090963
                                        ide/ptr-Insulinase family/Protease III-(CT806)
                                  R
CPn0958
           1094785
                       1093793
                                        plsB-Glycerol-3-P Acyltransferase-(CT807)
                                  R
CPh@959
           1096343
                       1094799
                                        cafE-Axial Filament Protein-(CT808)
                                  R
CPn:0960
           1096764
                       1097102
                                       CT809 hypothetical protein
CPn0961
           1097118
                       1097297
                                        rl32-L32 Ribosomal Protein-(CT810)
                                  F
CPN 0962
           1097316
                       1098275
                                  F
                                       plsX-FA/Phospholipid Synthesis Protein-(CT811)
CPn0963
           1098398
                       1103224
                                        pmp_21-Polymorphic Outer Membrane Protein D Family-(CT812)
                                  F
CPn0964
           1104758
                       1103301
                                  R
CPn0965
           1106736
                       1104925
                                        lpxB-Lipid A Disaccharide Synthase-(CT411)
                                  R
CPn0966
           1108037
                       1106748
                                        pcnB_2-PolyA Polymerase_2-(CT410)
                                  R
CPn0967
           1108512
                       1109885
                                  F
                                       mrsA/pgm-Phosphoglucomutase-(CT815)
CPn0968
           1109895
                       1111721
                                        glmS-Glucosamine-Fructose-6-P Aminotransferase-(CT816)
                                  F
CPn0969
           1111812
                       1112999
                                       0969-tyrP_1-Tyrosine Transport_1-(CT817) tyrP_1-Tyrosine Transport_1-
                                          (CT817)
CPn0970
           1113461
                                        0970-tyrP_2-Tyrosine Transport_2-(CT818) tyrP_2-Tyrosine Transport_2-
                      1114648
                                  F
                                          (CT818)
CPn0971
           1114702
                      1115415 .
                                  F
                                       yccA-Transport Permease-(CT819)
CPn0972
           1116299
                       1115430
                                  R
                                        ftsY-Cell Division Protein FtsY-(CT820)
CPn0973
                                       sucC-Succinyl-CoA Synthetase, Beta-(CT821)
           1116370
                       1117527
                                  F
CPn0974
           1117544
                       1118422
                                  F
                                        sucD-Succinyl-CoA Synthetase, Alpha-(CT822)
CPn0975
           1119104
                       1119637
                                  F
CPn0976
           1120082
                       1121185
                                  F
CPn0977
           1121371
                       1122402
                                  F
CPn0978
           1122665
                       1123693
                                  F
CPn0979
           1123980
                       1125443
                                  F
                                       htrA-DO Serine Protease-(CT823)
                                        *similarity to Saccharomyces serevisiae hypothetical 52.9KD protein
CPn0980
           1126982
                      1125504
                                  R
CPn0981
           1127031
                       1129952
                                       Zinc Metalloprotease (insulinase family) - (CT824)
                                  F
CPn0982
           1131194
                       1129962
                                  R
                                       yigN family-(CT825)
CPn0983
           1132000
                       1131206
                                       pssA-Glycerol-Serine Phosphatidyltransferase-(CT826)
                                  R
CPn0984
           1132379
                      1135510
                                  F
                                       nrdA-Ribonucleoside Reductase, Large Chain-(CT827)
CPn0985
           1135534
                      1136571
                                  F
                                       nrdB-Ribonucleoside Reductase, Small Chain-(CT828)
CPn0986
           1136724
                       1137395
                                       yggH-predicted rRNA Methylase-(CT829)
           1137516
                                       ytgB-like predicted rRNA methylase-(CT830)
CPn0987
                       1138115
                                  F
CPn0988
           1138986
                       1138075
                                  R
                                       murB-UDP-N-Acetylenolpyruvoylglucosamine Reductase-(CT831)
CPn0989
           1139495
                       1139016
                                  R
                                       CT832 hypothetical protein
CPn0990
           1139883
                       1140440
                                  F
                                       infC-Initiation Factor 3-(CT833)
CPn0991
           1140421
                       1140612
                                  F
                                       rl35-L35 Ribosomal Protein-(CT834)
```

```
rl20-L20 Ribosomal Protein-(CT835)
CPn0992
           1140634
                                       pheS-Phenylalanyl tRNA Synthetase, Alpha-(CT836)
           1141014
                      1142030
                                  F
CPn0993
                                       CT837 hypothetical protein
CPn0994
           1142398
                      1144440
                                  F
                      1144415
                                       CT838 hypothetical protein
           1145512
                                  R
CPn0995
                                       CT839 hypothetical protein
CPn0996
           1146589
                      1145519
                                       mesJ-PP-loop superfamily ATPase-(CT840)
                      1147664
                                  F
           1146708
CPn0997
                                       ftsH-ATP-dependent zinc protease-(CT841)
           1147855
                      1150584
                                  F
CPn0998
                                       pnp-Polyribonucleotide Nucleotidyltransferase-(CT842)
                      1150766
                                  R
           1152847
CPn0999
                                       rs15-S15 Ribosomal Protein-(CT843)
                      1152891
           1153157
                                  R
CPn1000
                                       yfhC-cytosine deaminase-(CT844)
           1153405
                      1153869
CPn1001
                                       CT845 hypothetical protein
                                  F
                      1154089
           1153862
CPn1002
                                       CT846 hypothetical protein
                      1154092
                                  R
           1154796
CPn1003
                                       CT847 hypothetical protein
                      1154879
                                  R
           1155397
CPn1004
                                       CT848 hypothetical protein
           1155933
                      1155415
                                  R
CPn1005
                                       CT849 hypothetical protein
           1156472
                      1155990
                                  R
CPn1006
                                       CT849.1 hypothetical protein
CPn1007
           1156689
                      1156907
                                  F
                                       CT850 hypothetical protein
                      1158223
                                  F
           1156928
CPn1008
                                       map-Methionine Aminopeptidase-(CT851)
CPn1009
           1159058
                      1158186
                                  R
                                       CT852 hypothetical protein
                      1159067
                                  R
CPn1010
           1159672
                                       CT853 hypothetical protein
           1160306
                      1159902
                                  R
CPn1011
                                       yzeB-ABC transporter permease-(CT854)
                      1160421
                                  R
           1162193
CPn1012
                                        fumC-Fumarate Hydratase-(CT855)
CPn1013
           1162245
                      1163624
                                  F
                                       ychM-Sulfate Transporter-(CT856)
           1165426
                      1163732
                                  R
CPn1014
                                       CT857 hypothetical protein (possible IM protein)
           1165634
                      1166893
                                  F
CPn:1015
                                       CT858 hypothetical protein
                      1168898
CPnf016
           1167042
                                  F
                                       lytB-Metalloprotease-(CT859)
           1169006
                      1169935
                                  F
CPn:1017
CPn1018
           1169898
                      1170629
                                  F
                                       CT860 hypothetical protein
                      1170638
CPn 1019
           1172128
                                  R
                                       CT861 hypothetical protein
           1173679
                      1172150
                                  R
CPn1020
                                       lcrH_2-Low Calcium Response_2-(CT862)
                      1173698
                                  R
CPn1021
           1174213
                                       CT863 hypothetical protein
           1175673
                                  R
                      1174216
CPn1022
           1176035
                      1176331
                                  F
CPn1023
                                       xerD-Integrase/recombinase-(CT864)
                      1176334
CPril 024
           1177236
                                  R
                                       pgi-Glucose-6-P Isomerase-(CT378)
                      1178879
CPn 1025
           1177302
                                        ltuA-(CT377)
                                  F
                      1179137
CPn1026
           1178997
CPn1027
           1179175
                      1180755
                                  F
                                       mdhC-Malate Dehyrogenase-(CT376)
CPn1028
           1181016
                      1181999
                                  F
           1182008
                      1182844
CPn 1029
                                       predicted D-amino acid dehyrogenase-(CT375)
CPn1030
           1183886
                      1182843
                                  R
                                        arcD-Arginine/Ornithine Antiporter-(CT374)
           1185552
                      1184098
                                  R
CPnF031
                                        CT373 hypothetical protein
                      1185566
CPnI032
           1186150
                                  R
                                        CT372 hypothetical protein
           1187500
                      1186187
CPn1033
                                  R
                                        Predicted OMP_1 (CT371) [leader (18) peptide]
CPn1034
           1188517
                      1187732
                                  R
                                        AroE-Shikimate 5-Dehyrogenase-(CT370)
CPn1035
           1190000
                      1188570
                                  R
                                        AroB-Dehyroquinate Synthase-(CT369)
           1191135
                      1189984
                                  R
CPn1036
                                        AroC-Chorismate Synthase-(CT368)
           1192199
                       1191123
                                  R
CPn1037
                                        aroL-Shikimate Kinase II-(CT367)
           1192726
                      1192199
CPn1038
                                  R
                                        aroA-Phosphoshikimate Vinyltransferase-(CT366)
           1193999
                       1192665
CPn1039
                                  R
CPn1040
                      1194073
           1194741
                                  R
                                        *bioA-Adenosylmethionine-8-Amino-7-Oxononanoate Aminotransferase
                       1194726
CPn1041
           1195994
                                  R
                                        *bioD-dethiobiotin synthetase
CPn1042
           .1196590
                      1195934
                                  R
                                        bioF_2-Oxononanoate Synthase_2
CPn1043
           1197717
                       1196572 .
                                  R
                                        *bioB-Biotin Synthase
CPn1044
           1198691
                       1197699
                                  R
                                        *conserved hypothetical bacterial membrane protein
CPn1045
           1199590
                       1198901
                                  R
                                        *Tryptophan Hyroxylase
CPn1046
           1200675
                       1199590
                                  R
                                        dapB-Dihydrodipicolinate Reductase-(CT364)
CPn1047
           1200552
                       1201343
                                        asd-Aspartate Dehydrogenase-(CT363)
CPn1048
           1201606
                       1202604
                                  F
                                        lysC-Aspartokinase III-(CT362)
CPn1049
           1202595
                       1203914
                                  F
                                        dapA-Dihydrodipicolinate Synthase-(CT361)
CPn1050
           1203926
                       1204798
                                  F
CPn1051
           1204962
                       1205270
                                  F
CPn1052
           1205417
                       1206169
                                  F
CPn1053
                       1206701
                                  F
           1206153
                                  F
CPn1054
           1207034
                       1209466
CPn1055
            1209694
                       1210521
                                  F
CPn1056
            1210527
                       1211228
                                  F
                                        CT356 hypothetical protein
           1211497
                       1213596
CPn1057
                                  F
                                        CT355 hypothetical protein
CPn1058
            1213748
                       1214836
                                  F
                                        kgsA-Dimethyladenosine Transferase-(CT354)
           1214848
CPn1059
                       1215678
                                  F
                                        dxs/tkt-Transketolase-(CT331)
CPn1060
            1217658
                       1215727
                                  R
                                        CT330 hypothetical protein
CPn1061
            1217920
                       1217666
                                  R
                                        xseA-Exodoxyribonuclease VII-(CT329)
CPn1062
            1219820
                       1218159
                                  R
                                        tpiS-Triosephosphate Isomerase-(CT328)
CPn1063
            1219951
                       1220712
```

1140996

F

```
CPn1064
          1220719
                     1220895
                                 F
CPn1065
          1221095
                      1220928
CPn1066
          1221135
                      1221488
                                 F
          1221735
                                      def-Polypeptide Deformylase-(CT353)
CPn1067
                      1222292
                                 F
CPn1068
          1223258
                     1222365
                                      rnhB_2-Ribonuclease HII_2-(CT008)
                                 R
CPn1069
          1223513
                     1223941
                                F
                                      yfgA-HTH Transcriptional Regulator-(CT009)
CPn1070
          1225511
                     1224144
                                 R
CPn1071
          1227324
                     1225885
                                R
CPn1072
          1227969
                     1228835
                                F
CPn1073
          1229011
                     1229832
                                F
                                      Predicted OMP_2 -(CT371)
```

Table 2 (Supplemental Data) Functional Assignments of *C. pneumoniae* Coding Sequences *C. trachomatis* genes are shown in parentheses

Amino Acid Biosynthesis

	9				Amino Acid biosynthesis
		Aromatic	Family		
		1039	(CT366)	aroA	Phosphoshikimate Vinyltransferase
		1036	(CT369)	aroB	Dehyroquinate Synthase
		1037	(CT368)	aroC	Chonsmate Synthase
	10	1035	(CT370)	aroÈ	Shikimate 5-Dehyrogenase
		0484	(CT382)	атоС	Deoxyheptonate Aldolase
		1038	(CT367)	aroL	Shikimate Kinase II
		0740	(CT637)	tyrB	Aroi vatic AA Aminotransferase
		Aspartate	e Family (lys	ine)	
	15	1048	(CT363)	asd	Aspartate Dehydrogenase
		1050	(CT361)	dapA	Dihydrodipicolinate Synthase
200		1047	(CT364)	dapB	Dihydrodipicolinate Reductase
LE.		0519	(CT430)	dapF	Dian inopimelate Epimerase
Title Title		1049	(CT362)	lysC	Aspa tokinase III
The state of the s	20	Serine Fa		•	•
# 17 T		0433	(CT282)	gcsH	Glyc ne Cleavage System H Protein
Rese		0521	(CT432)	glyA	Senne Hydroxymethyltransferase
					Some Hydroxymethyldansiciase
of the			lucleotide M		0.70
	25	0171		guaA	GMP Synthase
15 11 -	23	0172		guaB	Inosine 5'-Monophosphase Dehydrogenase
-		0608			Undine 5'-Monophosphate Synthase
i ni		0735	(00.00)		Undine Kinase
		0244	(CT128)	adk	Adenylate Kınase
	20	0894	(CT751)	amn	AMP Nucleosidase
	30	0568	(CT452)	cmk	CMP Kinase
		0392	(CT039)	dcd	dCTP Dearminase
		0059	(CT292)	dut	dUTP Nucleotidohydrolase
		0120	(CT030)	gmk	GMP Kinase
	25	0619	(CT500)	ndk	Nucleoside-2-P Kinase
	35	0984	(CT827)	nrdA	Ribonucleoside Reductase, Large Chain
		0985	(CT828)	nrdB	Ribonucieoside Reductase, Small Chain
		0236	(CT183)	рутС	CTP Synthetase
		0698	(CT678)	pyrH	UMP Kinase
	40	0273	(CT188)	tdk	Thymidylate Kinase
	40	0659	(CT539)	trxA	Thioredoxin
		0314	(CT099)	trxB	Thioredoxin Reductase
		1001	(CT844)	yfhC	Cytosine Deaminase
	15				Biosynthesis of Cofactors
	45		ipoate & Ubi	quinone	
		1041		bioA	Adenosylmethionine-8-Amino-7-Oxononanoate Aminotransferase
		1044		bioB	Biotin Synthase
		1042		bioD	Dethiobiotin Synthetase
	.	0923	(CT777)	bioF_1	Oxononanoate Synthase_I
	50	1043	(CT777)	bioF_2	Oxononanoate Synthase_2
		0866	(CT725)	bırA	Biotin Synthetase
		0748	(CT628)	ıspA	Geranyl Transtransferase
		0832	(CT558)	lıpA	Lipoate Synthetase

	0265	(CT219)	ubıA	Benzoate Octaphenyltransferase
	0264	(CT220)	ubıD	Phenylacrylate Decarboxylase
	0515	(CT428)	ubıE	Ubiquinone Methyltransferase
	Folic Acu	d		
5	0759	(CT612)	folA	Dihydrofolate Reductase
	0335	(CT078)	folD	Methylene Tetrahydrofolate Dehydrogenase
	0758	(CT613)	folP	Dihydropteroate Synthase
	0757	(CT614)	folX	Dihydroneopterin Aldolase
	0763	(CT649)	ygfA	Formyltetrahydrofolate Cycloligase
10		·	75171	1 omytedanyarototake Cyclongase
10	Porphyrin			
	0714	(CT662)	hemA	Glutamyl tRNA Reductase
	0744	(CT633)	hemB	Porphobilinogen Synthase
	0052	(CT299)	hemC	Porphobilinogen Deaminase
	0890	(CT747)	hemE	Uroporphyrinogen Decarboxylase
15	0888	(CT745)	hemG .	protoporphynnogen Oxidase
	0138	(CT210)	hemL	Glutamate-1-Semialdehyde-2,1-Aminomutase
	0380	(CT052)	$hemN_1$	Coproporphynnogen III Oxidase_1
	0889	(CT746)	hemN_2	Coproporphyrnnogen III Oxidase_2
	0603	(CT485)	hemZ	Ferrochetalase
20	Riboflavin	!		
	0872	(CT731)	пъА&пъВ	GTP Cyclohydratase & DHBP Synthase
	0532	(CT405)	пъС	Riboflavin Synthase
	0871	(CT730)	nbD	Riboflavin Deaminase
	0873	(CT732)	nbE	Ribityllumazine Synthase
25	0320	(CT093)	пbF	FAD Synthase
		,		2
	Cell Enve	lone		
		l & Phosphol	inid Metaha	dism
	0161	(CT206)	pia melabo	(predicted acyltransferase family)
30	0922	(CT776)	aas	
50	0414	(CT265)		Acylglycerophosphoethanolamine Acyltransferase
	0183		accA accB	AcCoA Carboxylase/Transferase Alpha
	0182	(CT123)		Biotin Carboxyl Carrier Protein
		(CT124)	accC	Biotin Carboxylase
35	0058	(CT293)	accD	AcCoA Carboxylase/Transferase Beta
55	0295	(CT236)	acpP	Acyl Carrier Protein
	0313	(CT100)	acpS	Acyl-carrier Protein Synthase
	0567	(CT451)	cdsA	Phosphatidate Cytidylytransferase
	0297	(CT238)	fabD	Malonyl Acyl Carner Transcyclase
40	0916	(CT770)	fabF	Acyl Carrier Protein Synthase
40	0296	(CT237)	fabG	Oxoacyl (Carrier Protein) Reductase
	0298	(CT239)	fabH	Oxoacyl Carrier Protein Synthase III
	0406	(CT104)	fabI	Enoyl-Acyl-Carrier Protein Reductase
	0651	(CT532)	fabZ	Mynstoyl-Acyl Carner Dehydratase
	0098	(CT010)	htrB	Acyltransferase
45	0271	(CT136)		Lysophospholipase Esterase
	0615	(CT496)	pgsA_l	Glycerol-3-P Phosphatidyltransferase_1
	0947	(CT797)	pgsA_2	Glycerol-3-P Phosphatydyltransferase_2
	0958	(CT807)	plsB	Glycerol-3-P Acyltransferase
	0569	(CT453)	pisC	Glycerol-3-P Acyltransferase
50	0962	(CT811)	plsX	FA/Phospholipid Synthesis Protein
	0839	(CT699)	psdD	Phosphatidylserine Decarboxylase
	0983	(CT826)	pssA	Glycerol-Serine Phosphatidyltransferase
	0921	(CT775)		snGlycerol-3-P Acyltransferase
	0654	(CT535)	yc1A	Acyl-CoA Thioesterase
55	0877	(CT736)	ybcL	CT736 Hypothetical Protein
-		,	. =	ME
	LPS			

gseA

KDO Transferase

0154

113

W

Į,Fi

ij

L.

i nii

į.

14

ũ

4

(CT208)

	0478	(CT379)	hflX	GTP Binding Protein
	0703	(CT673)		S/T Protein Kinase
	0095	(CT301)		S/T Protein Kinase
	0397	(CT259)		PP2C Phosphatase Family
5	0037	(CT337)	ptsH	PTS Phosphocarrier Protein Hpr
	0038	(CT336)	ptsI	PTS PEP Phosphotransferase
	0060	(CT291)	ptsN_1	PTS IIA Protein_1
	0061	(CT290)	ptsN_2	PTS IIA Protein + HTH DNA-Binding Domain
	0262	(CT218)	surE	SurE-like Acid Phosphatase
10	0838	(CT698)	thdF	Thiophene/Furan Oxidation Protein
	0693	(CT683)		TPR Repeats-CT683 Hypothetical Protein
	0321	(CT092)	ychF	GTP Binding Protein
	0544	(CT418)	yhbZ	GTP binding protein
	0844	(CT703)	yphC	GTPase/GTP-binding protein
15	Standard	Protein Seci	retion	
	0115	(CT025)	ffh	Signal Recognition Particle GTPase
	0363	(CT060)	flhA	Flagellar Secretion Protein
	0858	(CT717)	fliI	Flagellum-specific ATP Synthase
	0704	(CT672)	fliN	Flagellar Motor Switch Domain/YscQ family
20	0815	(CT572)	gspD	Gen. Secretion Protein D
	0816	(CT571)	gspE	Gen. Secretion Protein E
	0817	(CT570)	gspF	Gen. Secretion Protein F
	0359	(CT064)	lepA	GTPase
	0110	(CT020)	lepB	Signal Peptidase I
25	0535	(CT408)	IspA	Lipoprotein Signal Peptidase
	0260	(CT141)	secA_1	Protein Translocase Subunit_1
	0841	(CT701)	secA_2	Translocase SecA_2
	0564	(CT448)	secD&sec	F Protein Export Proteins SecD/SecF (fusion)
20	0075	(CT321)	secĒ	Preprotein Transiocase
30	0629	(CT510)	secY	Translocase
	0848	(CT707)	tig	Tngger Factor-Peptidyl-prolyl Isomerase
		t-Related Pr	oteins	
	0486			Hypothetical Proline Permease
35	0289	(CT230)	aaaT	Neutral Amino Acid (Glutamate) Transporter
33	0691	(CT685)	abcX	ABC Transporter ATPase
	1031	(CT374)	arcD	
				Arginine/Ornithine Antiporter
	0482	(CT381)	artJ	Arginine Penplasmic Binding Protein
	0836	(CT381) (CT554)	artJ brnQ	Arginine Periplasmic Binding Protein Amino Acid (Branched) Transport
40	0836 0536	(CT381) (CT554) (CT409)	artJ brnQ dagA_1	Arginine Periplasmic Binding Protein Amino Acid (Branched) Transport D-Ala/Gly Permease_1
40	0836 0536 0876	(CT381) (CT554) (CT409) (CT735)	artJ brnQ dagA_1 dagA_2	Arginine Periplasmic Binding Protein Amino Acid (Branched) Transport D-Ala/Gly Permease_1 D-Alanine/Glycine Permease_2
40	0836 0536 0876 0682	(CT381) (CT554) (CT409) (CT735) (CT690)	artJ brnQ dagA_1 dagA_2 dppD	Arginine Periplasmic Binding Protein Amino Acid (Branched) Transport D-Ala/Gly Permease_1 D-Alanine/Glycine Permease_2 ABC ATPase Dipeptide Transport
40	0836 0536 0876 0682 0683	(CT381) (CT554) (CT409) (CT735) (CT690) (CT689)	art/ brnQ dagA_1 dagA_2 dppD dppF	Arginine Periplasmic Binding Protein Amino Acid (Branched) Transport D-Ala/Gly Permease_1 D-Alanine/Glycine Permease_2 ABC ATPase Dipeptide Transport ABC ATPase Dipeptide Transport
40	0836 0536 0876 0682 0683 0280	(CT381) (CT554) (CT409) (CT735) (CT690)	artJ bmQ dagA_1 dagA_2 dppD dppF dppF	Arginine Periplasmic Binding Protein Amino Acid (Branched) Transport D-Ala/Gly Permease_1 D-Alanine/Glycine Permease_2 ABC ATPase Dipeptide Transport ABC ATPase Dipeptide Transport Dipeptide Transporter ATPase
40	0836 0536 0876 0682 0683	(CT381) (CT554) (CT409) (CT735) (CT690) (CT689) (CT689) (CT596)	art/ brnQ dagA_1 dagA_2 dppD dppF	Arginine Periplasmic Binding Protein Amino Acid (Branched) Transport D-Ala/Gly Permease_1 D-Alanine/Glycine Permease_2 ABC ATPase Dipeptide Transport ABC ATPase Dipeptide Transport
	0836 0536 0876 0682 0683 0280 0785	(CT381) (CT554) (CT409) (CT735) (CT690) (CT689) (CT689)	artJ brnQ dagA_1 dagA_2 dppD dppF dppF exbB	Arginine Periplasmic Binding Protein Amino Acid (Branched) Transport D-Ala/Gly Permease_1 D-Alanine/Glycine Permease_2 ABC ATPase Dipeptide Transport ABC ATPase Dipeptide Transport Dipeptide Transporter ATPase Macromolecule Transporter
	0836 0536 0876 0682 0683 0280 0785	(CT381) (CT554) (CT409) (CT735) (CT690) (CT689) (CT689) (CT596) (CT597)	artJ brnQ dagA_1 dagA_2 dppD dppF dppF exbB exbD	Arginine Periplasmic Binding Protein Amino Acid (Branched) Transport D-Ala/Gly Permease_1 D-Alanine/Glycine Permease_2 ABC ATPase Dipeptide Transport ABC ATPase Dipeptide Transport Dipeptide Transporter ATPase Macromolecule Transporter Biopolymer Transport Protein
	0836 0536 0876 0682 0683 0280 0785 0784	(CT381) (CT554) (CT409) (CT735) (CT690) (CT689) (CT689) (CT596) (CT597) (CT486)	artJ brnQ dagA_1 dagA_2 dppD dppF dppF exbB exbD	Arginine Periplasmic Binding Protein Amino Acid (Branched) Transport D-Ala/Gly Permease_1 D-Alanine/Glycine Permease_2 ABC ATPase Dipeptide Transport ABC ATPase Dipeptide Transport Dipeptide Transporter ATPase Macromolecule Transporter Biopolymer Transport Protein Glutamine Binding Protein
	0836 0536 0876 0682 0683 0280 0785 0784 0604	(CT381) (CT554) (CT409) (CT735) (CT690) (CT689) (CT689) (CT596) (CT597) (CT486) (CT129)	artJ brnQ dagA_1 dagA_2 dppD dppF dppF exbB exbD fliY glnP	Arginine Periplasmic Binding Protein Amino Acid (Branched) Transport D-Ala/Gly Permease_1 D-Alanine/Glycine Permease_2 ABC ATPase Dipeptide Transport ABC ATPase Dipeptide Transport Dipeptide Transporter ATPase Macromolecule Transporter Biopolymer Transport Protein Glutamine Binding Protein ABC Amino Acid Transporter Permease
45	0836 0536 0876 0682 0683 0280 0785 0784 0604 0192 0191	(CT381) (CT554) (CT409) (CT735) (CT690) (CT689) (CT689) (CT596) (CT597) (CT486) (CT129) (CT130) (CT401)	artJ bmQ dagA_1 dagA_2 dppD dppF dppF exbB exbD fliY glnP glnQ gltT	Arginine Periplasmic Binding Protein Amino Acid (Branched) Transport D-Ala/Gly Permease_1 D-Alanine/Glycine Permease_2 ABC ATPase Dipeptide Transport ABC ATPase Dipeptide Transport Dipeptide Transporter ATPase Macromolecule Transporter Biopolymer Transport Protein Glutamine Binding Protein ABC Amino Acid Transporter ATPase Glutamate Symport
	0836 0536 0876 0682 0683 0280 0785 0784 0604 0192 0191 0528	(CT381) (CT554) (CT409) (CT735) (CT690) (CT689) (CT689) (CT596) (CT597) (CT486) (CT129) (CT130) (CT401)	artJ brnQ dagA_1 dagA_2 dppD dppF dppF exbB exbD fliY glnP glnQ	Arginine Periplasmic Binding Protein Amino Acid (Branched) Transport D-Ala/Gly Permease_1 D-Alanine/Glycine Permease_2 ABC ATPase Dipeptide Transport ABC ATPase Dipeptide Transport Dipeptide Transporter ATPase Macromolecule Transporter Biopolymer Transport Protein Glutamine Binding Protein ABC Amino Acid Transporter Permease ABC Amino Acid Transporter ATPase Glutamate Symport Mg++ Transporter (CBS Domain)
45	0836 0536 0876 0682 0683 0280 0785 0784 0604 0192 0191 0528 0286 0413	(CT381) (CT554) (CT409) (CT409) (CT735) (CT690) (CT689) (CT689) (CT596) (CT597) (CT486) (CT129) (CT130) (CT401) (CT194) (CT194)	artJ bmQ dagA_1 dagA_2 dppD dppF dppF exbB exbD fliY glnP glnQ gltT mgtE	Arginine Periplasmic Binding Protein Amino Acid (Branched) Transport D-Ala/Gly Permease_1 D-Alanine/Glycine Permease_2 ABC ATPase Dipeptide Transport ABC ATPase Dipeptide Transport Dipeptide Transporter ATPase Macromolecule Transporter Biopolymer Transport Protein Glutamine Binding Protein ABC Amino Acid Transporter Permease ABC Amino Acid Transporter ATPase Glutamate Symport Mg ⁺⁺ Transporter (CBS Domain) Transport ATP Binding Protein
45	0836 0536 0876 0682 0683 0280 0785 0784 0604 0192 0191 0528 0286 0413	(CT381) (CT554) (CT409) (CT409) (CT735) (CT690) (CT689) (CT689) (CT596) (CT597) (CT486) (CT129) (CT130) (CT401) (CT194) (CT264)	artJ brmQ dagA_1 dagA_2 dppD dppF dppF exbB exbD fliY glnP glnQ gltT mgtE msbA	Arginine Periplasmic Binding Protein Amino Acid (Branched) Transport D-Ala/Gly Permease_1 D-Alanine/Glycine Permease_2 ABC ATPase Dipeptide Transport ABC ATPase Dipeptide Transport Dipeptide Transporter ATPase Macromolecule Transporter Biopolymer Transport Protein Glutamine Binding Protein ABC Amino Acid Transporter ATPase Glutamate Symport Mg++ Transporter (CBS Domain) Transport ATP Binding Protein Na+-dependent Transporter
45	0836 0536 0876 0682 0683 0280 0785 0784 0604 0192 0191 0528 0286 0413	(CT381) (CT554) (CT409) (CT735) (CT690) (CT689) (CT689) (CT596) (CT597) (CT486) (CT129) (CT130) (CT401) (CT194) (CT264) (CT231) (CT198)	artJ bmQ dagA_1 dagA_2 dppD dppF dppF exbB exbD fliY glnP glnQ gltT mgtE msbA	Arginine Periplasmic Binding Protein Amino Acid (Branched) Transport D-Ala/Gly Permease_1 D-Alanine/Glycine Permease_2 ABC ATPase Dipeptide Transport ABC ATPase Dipeptide Transport Dipeptide Transporter ATPase Macromolecule Transporter Biopolymer Transport Protein Glutamine Binding Protein ABC Amino Acid Transporter ATPase Glutamate Symport Mg++ Transporter (CBS Domain) Transport ATP Binding Protein Na+-dependent Transporter Oligopeptide Binding Protein_1
45	0836 0536 0876 0682 0683 0280 0785 0784 0604 0192 0191 0528 0286 0413 0290 0195	(CT381) (CT554) (CT409) (CT409) (CT735) (CT690) (CT689) (CT596) (CT597) (CT486) (CT129) (CT130) (CT401) (CT194) (CT264) (CT231) (CT198)	artJ bmQ dagA_1 dagA_2 dppD dppF dppF exbB exbD fliY glnP glnQ gltT mgtE msbA	Arginine Periplasmic Binding Protein Amino Acid (Branched) Transport D-Ala/Gly Permease_1 D-Alanine/Glycine Permease_2 ABC ATPase Dipeptide Transport ABC ATPase Dipeptide Transport Dipeptide Transporter ATPase Macromolecule Transporter Biopolymer Transport Protein Glutamine Binding Protein ABC Amino Acid Transporter Permease ABC Amino Acid Transporter ATPase Glutamate Symport Mg ⁺⁺ Transporter (CBS Domain) Transport ATP Binding Protein Na ⁺ -dependent Transporter Oligopeptide Binding Protein_1 Oligopeptide Binding Protein_2
45	0836 0536 0876 0682 0683 0280 0785 0784 0604 0192 0191 0528 0286 0413	(CT381) (CT554) (CT409) (CT735) (CT690) (CT689) (CT689) (CT596) (CT597) (CT486) (CT129) (CT130) (CT401) (CT194) (CT264) (CT231) (CT198)	artJ bmQ dagA_1 dagA_2 dppD dppF dppF exbB exbD fliY glnP glnQ gltT mgtE msbA	Arginine Periplasmic Binding Protein Amino Acid (Branched) Transport D-Ala/Gly Permease_1 D-Alanine/Glycine Permease_2 ABC ATPase Dipeptide Transport ABC ATPase Dipeptide Transport Dipeptide Transporter ATPase Macromolecule Transporter Biopolymer Transport Protein Glutamine Binding Protein ABC Amino Acid Transporter ATPase Glutamate Symport Mg++ Transporter (CBS Domain) Transport ATP Binding Protein Na+-dependent Transporter Oligopeptide Binding Protein_1

Oligopeptide Binding Lipoprotein_5

0599

ļ.,

H

i di

ļ.

į "i

-

.T

57

(CT480)

oppA_5

```
(CT087)
                                    malQ
                  0851
                         (CT710)
                                    pck,A
                                              Phosphoenolpyruvate Carboxykinase
               Phosphorous & Sulfur
                  0548
                         (CT435)
                                              Sulfite Reductase
                                    cysJ
        5
                  0920
                         (CT774)
                                    cysQ
                                              Sulfite Synthesis/Biphosphate Phosphatase
                  0025
                         (CT346)
                                              Sulphohydrolase
                                    atsA
                  0918
                        (CT772)
                                              Inorganic Pyrophosphatase
                                    ppa
                                                        DNA Replication, Modification, Repair & Recombination
      10
               DNA Mismatch Repair
                  0505
                                              3-Methyladenine DNA Glycosylase
                  0812
                         (CT575)
                                    mutL
                                              DNA Mismatch Repair
                  0941
                         (CT792)
                                    mutS
                                               DNA Mismatch Repair
                  0402
                         (CT107)
                                    mutY
                                               Adenine Glycosylase
      15
                  0732
                         (CT625)
                                    nfo
                                               Endonuclease IV
                  0837
                         (CT697)
                                               Enodnuclease III
                                    nth
               DNA Modification
                  0596
                         (CT477)
                                    ada
                                              Methyltransferase
                  0114
                         (CT024)
                                    hemK
                                              A/G-specific Methylase
THE SALE SALE SALES
      20
                  0891
                         (CT748)
                                    mfd
                                              Transcription-Repair Coupling
                  0620
                         (CT501)
                                              Holliday Junction Helicase
                                    ruvA
                  0390
                         (CT040)
                                    ruvB
                                              Holliday Junction Helicase
                         (CT502)
                                              Crossover Junction Endonuclease
                  0621
                                    ruvC
                  0053
                         (CT298)
                                              Sms Protein
                                    sms
į,
      25
                  0773
                         (CT607)
                                    ung
                                              Uracil DNA Glycosylase
ij
                  1062
                         (CT329)
                                               Exodoxymbonuclease VII
                                    xseA
DNA Recombination
=
                  0762
                         (CT650)
                                              RecA Recombination Protein
                                    гесА
Į.i.
                  0738
                         (CT639)
                                    recB
                                               Exodeoxynbonuclease V, Beta
30
                  0737
                         (CT640)
                                    recC
                                               Exodeoxynbonuclease V, Gamma
14
                  0123
                         (CT033)
                                    recD_1
                                               Exodeoxynbonuclease V (Alpha Subunit)_1
-
                  0752
                         (CT652)
                                    recD_2
                                              Exodeoxynbonuclease V, Alpha_2
M
                  0339
                         (CT074)
                                               ABC Superfamily ATPase
                                    recF
1
                  0340
                         (CT074)
                                              (frame-shift with 0339)
      35
                  0563
                         (CT447)
                                               ssDNA Exonuclease
                                    recJ
                  0299
                         (CT240)
                                    recR
                                               Recombination Protein
               DNA Replication
                  0309
                         (CT250)
                                    dnaA_1
                                              Replication Initiation Protein_l
                  0424
                         (CT275)
                                    dnaA_2
                                               Replication Initiation Factor_2
      40
                  0616
                         (CT497)
                                    dnaB
                                               Replicative DNA Helicase
                  0666
                         (CT545)
                                    dnaE
                                               DNA Pol III Alpha
                                               DNA Primase
                  0942
                         (CT794)
                                    dnaG
                  0338
                         (CT075)
                                    dnaN
                                               DNA Poi III (Beta)
                  0410
                         (CT261)
                                    dnaQ_1
                                              DNA Pol III Epsilon Chain_l
      45
                         (CT536)
                  0655
                                    dnaQ_2
                                              DNA Pol III Epsilon Chain_2
                  0040
                         (CT334)
                                    dnaX_1
                                               DNA Pol III Gamma and Tau_I
                  0272
                         (CT187)
                                    dnaX_2
                                               DNA Pol III Gamma and Tau_2
                  0149
                         (CT146)
                                    dnlJ
                                               DNA Ligase
                  0274
                         (CT189)
                                    gytA_l
                                               DNA Gyrase Subunit A_I
      50
                  0716
                         (CT660)
                                    gyrA_2
                                               DNA Gyrase Subunit A_2
                                    gyrB_1
                  0275
                         (CT190)
                                               DNA Gyrase Subunit B_1
                  0715
                         (CT661)
                                     gyrB_2
                                               DNA Gyrase Subunit B_2
                  0416
                         (CT267)
                                    himD
                                               Integration Host Factor Alpha
                  0612
                         (CT493)
                                    polA
                                               DNA Polymerase I
      55
                  0924
                         (CT778)
                                    рπА
                                               Primosomal Protein N'
                  0386
                         (CT044)
                                               SS DNA Binding Protein
                                    ssb
```

Glucanotransferase

```
SWI/SNF family helicase_1
            0835
                  (CT555)
            0849
                  (CT708)
                                        SWI/SNF family helicase_2
                                        DNA Topoisomerase I-Fused to SWI Domain
            0769
                  (CT643)
                              topA
                  (CT347)
                                        Integrase/recombinase
            0024
                              xerC
 5
            1024
                  (CT864)
                              xerD
                                        Integrase/recombinase
         Eukaryotic-Type Chromatin Factors
                                        Histone-Like Developmental Protein
            0886
                  (CT743)
                              hctA
                                        Histone-like Protein 2
            0384
                  (CT046)
                              hctB
            0878
                  (CT737)
                                        SET Domain protein
10
                                        SWIB (YM74) Complex Protein
            0577
                  (CT460)
         UVR Exinuclease Repair System
            0096
                  (CT333)
                              uvtA
                                        Excinuclease ABC Subunit A
            0801
                  (CT586)
                              ичтВ
                                        Exinuclease ABC Subunit B
            0940
                  (CT791)
                              uvrC
                                        Excinuclease ABC, Subunit C
15
                                        DNA Helicase
            0772
                  (CT608)
                              uvrD
                                                                    Energy Metabolism
         Aerobic
                                        Glycerol-3-P Dehydrogenase
            0855
                  (CT714)
                              gpdA
20
                                        Ubiquinone Oxidoreductase, Alpha
            0743
                  (CT634)
                              nqrA
                              nqr2
                                        NADH (Ubiquinone) Dehydrogenase
            0427
                  (CT278)
                                        NADH (Ubiquinone) Oxidoreductase, Gamma
            0428
                  (CT279)
                              nqr3
                   (CT280)
                                        NADH (Ubiquinone) Reductase 4
            0429
                              nqr4
            0430
                  (CT281)
                                        NADH (Ubiquinone) Reductase 5
                              nar5
25
                                        Phenolhydrolase/NADH (Ubiquinone) Oxidoreductase 6
            0883
                  (CT740)
                              ngr6
       . ATP Biogenesis and metabolism
            0351
                  (CT065)
                              adt_l
                                        ADP/ATP Translocase_1
            0614
                  (CT495)
                              adt_2
                                        ADP/ATP Translocase 2
            0088
                   (CT308)
                                        ATP Synthase Subunit A
                              atpA
30
            0089
                   (CT307)
                              atpB
                                        ATP Synthase Subunit B
            0090
                   (CT306)
                              atpD
                                        ATP Synthase Subunit D
            0086
                   (CT310)
                              афЕ
                                        ATP Synthase Subunit E
            0091
                   (CT305)
                                        ATP Synthase Subunit I
                              atpl
            0092
                   (CT304)
                              аtрК
                                        ATP Synthase Subunit K
35
                  (CT719)
                                        Flagellar M-Ring Protein
            0860
                              flıF
         Electron Transport Chain
            0102
                   (CT013)
                                        Cytochrome Oxidase Subunit I
                              cydA
            0103
                   (CT014)
                                        Cytochrome Oxidase Subunit II
                              cydB
            0364
                   (CT059)
                                        Ferredoxin
40
            0084
                  (CT312)
                                        Predicted Ferredoxin
         Glycolysis & Gluconeogenesis
                                        Predicted 1,6-Fructose Biphosphate Aldolase
            0281
                   (CT215)
                              dhnA
                                         Enolase
            0800
                   (CT587)
                              eno
                   (CT505)
                                        Glyceraldehyde-3-P Dehyrogenase
            0624
                              gapA
45
            0056
                   (CT295)
                                        Phosphomannomutase
                              mrsA
                                         Phosphoglucomutase
            0967
                   (CT815)
                              pgm
            0160
                   (CT207)
                                        Fructose-6-P Phosphotransferase_1
                              pfkA_1
            0208
                   (CT205)
                              pfkA_2
                                        Fructose-6-P Phosphotransferase 2
                                         Glucose-6-P Isomerase
            1025
                   (CT378)
                              pgı
50
            0679
                   (CT693)
                              pgk
                                         Phosphoglycerate Kinase
            0863
                   (CT722)
                              pgmA
                                         Phosphoglycerate Mutase
            0097
                   (CT332)
                              pyk
                                         Pyruvate Kinase
            1063
                   (CT328)
                              φıS
                                         Triosephosphate Isomerase
          Pentose Phosphate Pathway
55
                                         Glucose-6-P Dehyrogenase (DevB family)
            0239
                   (CT186)
                              devB
                                         Transketolase
```

W

W

ļ.

D

1-1

Į-

4D

ıD

1060

(CT331)

dxs

6-Phosphogluconate Dehydrogenase

Ribulose-P Epimerase

0360

0185

55

0385

0136

(CT045)

(CT112)

pepA

pepF

(CT063)

(CT121)

gnd

трe

Leucyl Aminopeptidase A

Oligopeptidase

	0813	(CI5/4)	pepP	Aminopeptidase P
	0613	(CT494)	sohB	Protease
	0555	(CT441)	tsp	Tail-Specific Protease
	0344	(CT072)	yaeL	Metalloprotease
5	0981	(CT824)		Zinc Metalloprotease (insulinase family)
	Protein Is	omerases		
	0227	(CT176)	dsbB	Disulfide bond Oxidoreductase
	0786	(CT595)	dsbD	Thio disulfide Interchange Protein
	0228	(CT177)	dsbG	Disulfide Bond Chaperone
10	0933	(CT783)		Predicted Disulfide Bond Isomerase
	0926	(CT780)		Thioredoxin Disulfide Isomerase

Transcription

```
0999
                  (CT842)
                                        Polyribonucleotide Nucleotidyltransferase
                              pnp
 5
            0054
                  (CT297)
                              mc
                                        Ribonuclease III
                                        Ribonuclease HII_1
            0119
                  (CT029)
                              mhB_1
                                        Ribonuclease HII 2
            1068
                  (CT008)
                              mhB 2
                  (CT784)
                                        Ribonuclease P Protein Component
            0934
                              mpA
            0504
                  (CT397)
                              vacB
                                        Ribonuclease Family
10
         RNA Elongation & Termination Factors
                              greA
                                        Transcription Elongation Factor
            0741
                  (CT636)
                                        N Utilization Protein A
            0316
                  (CT097)
                              nusA
            0076
                   (CT320)
                              nusG
                                        Transcriptional Antitermination
                  (CT704)
                              pcnB 1
                                        Poly A Polymerase 1
            0845
15
                  (CT410)
                              pcnB_2
                                        PolyA Polymerase_2
            0966
            0610
                  (CT491)
                                        Transcription Termination Factor
         RNA Methylases
            0674
                  (CT553)
                                        RNA Methyltransferase
                              fmu
            1059
                  (CT354)
                                        Dimethyladenosine Transferase
                              kgsA
20
            0187
                  (CT133)
                                        Predicted Methylase
                  (CT403)
            0530
                              spoU_1
                                        rRNA Methylase_1
            0660
                  (CT540)
                              spoU_2
                                        rRNA Methylase_2
            0117
                  (CT027)
                              trmD
                                        tRNA (Guanine N-1)-Methyltransferase
            0885
                  (CT742)
                                        rRNA Methyltransferse
                              ygcA
25
            0986
                  (CT829)
                                        Predicted rRNA Methylase
                              yggH
            0987
                  (CT830)
                                        Predicted rRNA Methylase
                              ytgB
         RNA Modification
            0649
                  (CT530)
                              fmt
                                        Methionyl tRNA Formyltransferase
            0910
                  (CT766)
                                        tRNA Pyrophosphate Transferase
                              mıaA
30
            0719
                  (CT658)
                              sfhB
                                        Predicted Pseudoundine Synthase
            0219
                  (CT193)
                              tgt
                                        Queuine tRNA Ribosyl Transferase
            0580
                  (CT463)
                              truA
                                        Pseudoundylate Synthase I
            0319
                  (CT094)
                              truB
                                        tRNA Pseudoundine Synthase
            0403
                  (CT106)
                              yceC
                                        Predicted Pseudoundine Synthetase Family
35
           0864
                  (CT723)
                              ујЬС
                                        Predicted Pseudoundine Synthase
         RNA Polymerase & Transcription Regulators
            0586
                  (CT468)
                              atoC
                                        Two-Component Regulator
                                        Sigma-28/WhiG Family
            0362
                  (CT061)
                              rpsD
            0501
                  (CT394)
                              hrcA
                                        HTH Transcriptional Repressor
40
            0793
                  (CT588)
                                        Sigma Regulatory Family Protein-PP2C Phosphatase (RsbW Antagonist)
                              rbsU
            0626
                  (CT507)
                              гроА
                                        RNA Polymerase Alpha
            0081
                  (CT315)
                              гроВ
                                        RNA Polymerase Beta
            0082
                  (CT314)
                              гроС
                                        RNA Polymerase Beta'
            0756
                  (CT615)
                                        RNA Polymerase Sigma-66
                              rpoD
45
            0771
                  (CT609)
                              гроЙ
                                        RNA Polymerase Sigma-54
            0511
                  (CT424)
                              rsbV_1
                                        Sigma Regulatory Factor_I
            0909
                  (CT765)
                              rsbV_2
                                        Sigma Factor Regulator_2
            0670
                  (CT549)
                              rsbW
                                        Sigma Regulatory Factor-Histidine Kinase
                  (CT630)
            0750
                              tctD
                                        HTH Transcriptional Regulatory Protein + Receiver Doman
50
                  (CT009)
            1069
                                        HTH Transcriptional Regulator
                              yfgA
                                                                        Translation
         Amino Acyl tRNA Synthesis
            0892
                   (CT749)
                               alaS
                                        Alanyl tRNA Synthetase
55
            0570
                   (CT454)
                              argS
                                        Arginyl tRNA Transferase
            0662
                  (CT542)
                                        Aspartyl tRNA Synthetase
                              aspS
```

RNA Degradation

Hall falls and

113

ļ.i.

m

į...į.

14

H

T)

ų.

æ

	0932	(CT782)	cysS	Cysteinyl tRNA Synthetase
	0003	(CT003)	gatA	Glu tRNA Gln Amidotransferase (A subunit)
	0004	(CT004)	gatB	Glu tRNA Gln Amidotransferase (B Subunit)
	0002	(CT002)	gatC	Glu tRNA Gln Amidotransferase (C subunit)
5	0560	(CT445)	gitX	Glutarnyl-tRNA Synthetase
	0946	(CT796)	glyQ	Glycyl tRNA Synthetase
	0663	(CT543)	hisS	Histidyl tRNA Synthetase
	0109	(CT019)	ıleS	Isoleucyl-tRNA Synthetase
	0153	(CT209)	leuS	Leucyl tRNA Synthetase
10	0931	(CT781)	lysS	Lysyl tRNA Synthetase
	0122	(CT032)	metG	Methionyl-tRNA Synthetase
	0993	(CT836)	pheS	Phenylalanyi tRNA Synthetase, Alpha
	0594	(CT475)	pheT	Phenylalanyi tRNA Synthetase Beta
	0500	(CT393)	proS	Prolyi tRNA Synthetase
15	0870	(CT729)	serS	Seryi tRNA Synthetase_2
	0806	(CT581)	thrS	Threonyl tRNA Synthetase
	0802	(CT585)	trpS	Tryptophanyl tRNA Synthetase
	0361	(CT062)	tytS	Tyrosyl tRNA Synthetase
	0094	(CT302)	valS	Valyl tRNA Synthetase
20	Pepude C	hain Initiati	on, Elongai	tion & Termination
	1067	(CT353)	def	Polypeptide Deformylase
	0184	(CT122)	efp_1	Elongation Factor P_1
	0895	(CT752)	efp_2	Elongation Factor P_2
	0550	(CT437)	fusA	Elongation Factor G
25	0073	(CT323)	ınfA	Institution Factor IF-1
	0317	(CT096)	ınfB	Initiation Factor-2
	0990	(CT833)	ınfC	Initiation Factor 3
	0113	(CT023)	pfrA	Peptide Chain Releasing Factor I
	0576	(CT459)	prfB	Peptide Chain Release Factor 2
30	0950	(CT800)	pth	Peptidyl tRNA Hydrolase
	0318	(CT095)	rbfA	Ribosome Binding Factor A
	0699	(CT677)	пſ	Ribosome Releasing Factor
	0697	(CT679)	tsf	Elongation Factor TS
	0074	(CT322)	tufA	Elongation Factor Tu
35	Ribosoma	il Proteins		
	0078	(CT318)	ri I	L1 Ribosomal Protein
	0644	(CT525)	rl2	L2 Ribosomal Protein
	0647	(CT528)	ri3	L3 Ribosomal Protein
	0646	(CT527)	rl4	L4 Ribosomal Protein
40	0635	(CT516)	rl5	L5 Ribosomal Protein
	0633	(CT514)	rl6	L6 Ribosomal Protein
	0080	(CT316)	rl7	L7/L12 Ribosomal Protein
	0953	(CT803)	ri9	L9 Ribosomal Protein
	0079	(CT317)	rl10	L10 Ribosomal Protein
45	0077	(CT319)	rl I I	L11 Ribosomal Protein
	0247	(CT125)	rl13	L13 Ribosomal Protein
	0637	(CT518)	rl14	L14 Ribosomal Protein
	0630	(CT511)	rl15	L15 Ribosomal Protein
50	0640	(CT521)	rl16	L16 Ribosomal Protein
50	0625	(CT506)	ri17	L17 Ribosomal Protein
	0632	(CT513)	rl18	L18 Ribosomal Protein
	0118	(CT028)	rl19	L19 Ribosomal Protein
	0992	(CT835)	rl20	L20 Ribosomal Protein
55	0546	(CT420)	r121	L21 Ribosomal Protein
55	0642	(CT523)	rl22	L22 Ribosomal Protein
	0645	(CT526)	rl23	L23 Ribosomal Protein

		0636	(CT517)	rl24	L24 Ribosomal Protein
		0545	(CT419)	rl27	L27 ribosomal protein
		0327	(CT086)	ri28	L28 Ribosomal Protein
		0639	(CT520)	rl29	L29 Ribosomal Protein
	5	0112	(CT022)	rì31	L31 Ribosomal Protein
		0961	(CT810)	ri32	L32 Ribosomal Protein
		0250	(CT150)	ri33	L33 Ribosomal Protein
		0935	(CT785)	rl34	L34 Ribosomal Protein
		0991	(CT834)	rl35	L35 Ribosomal Protein
	10	0936	(CT786)	ri36	L36 Ribosomal Protein
		0315	(CT098)	rs l	S1 Ribosomal Protein
		0696	(CT680)	rs2	S2 Ribosomal Protein
		0641	(CT522)	rs3	S3 Ribosomal Protein
		0733	(CT626)	rs4	S4 Ribosomal Protein
	15	0631	(CT512)	rs5	S5 Ribosomal Protein
	13	0951	(CT801)	rs6	S6 Ribosomal Protein
					S7 Ribosomai Protein
		0551	(CT515)	rs7	
		0634	(CT515)	rs8	S8 Ribosomal Protein
	20	0246	(CT126)	159	S9 Ribosomal Protein
	20	0549	(CT436)	rs10	S10 Ribosomal Protein
ŧij.		0627	(CT508)	rs I I	S11 Ribosomal Protein
ion For		0552	(CT439)	rs12	S12 Ribosomal Protein
Appendix of the second		0628	(CT509)	rs13	S13 Ribosomal Protein
m	25	0937	(CT787)	rs14	S14 Ribosomal Protein
E al	25	1000	(CT843)	rs15	S15 Ribosomal Protein
		0116	(CT026)	rs16	S16 Ribosomal Protein
era Era		0638	(CT519)	rs17	S17 Ribosomal Protein
		0952	(CT802)	rs18	S18 Ribosomal Protein
E		0643	(CT524)	rs19	S19 Ribosomal Protein
ļ=	30	0754	(CT617)	rs20	S20 Ribosomal Protein
ļ.		0031	(CT342)	rs21	S21 Ribosomal Protein
1-					
4D					
H.	35				Other Categories
		Chlamydi	a-Specific P	Proteins	
		0561	(CT446)	Euo	CHLPS Euo Protein
		0804	(CT583)	Gp6D	CHLTR Plasmid Paralog
		0186	(CT119)		Similarity to IncA_I
	40	0291	(CT232)	ıncB	Inclusion Membrane Protein B
		0292	(CT233)	ıncC	Inclusion Membrane Protein C
		1026	(CT377)		LtuA Protein
		0333	(CT080)		LtuB Protein
		0005	(CT871)	pmp_1	Polymorphic Outer Membrane Protein G Family
	45	0013	(CT871)	pmp_2	Polymorphic Outer Membrane Protein G Family
		0014	(CT871)	pmp_3	Polymorphic Outer Membrane Protein G Family
		0015	(CT871)	pmp_3	PMP_3 (frame-shift with 0014)
		0016	(CT874)	pmp_4	Polymorphic Outer Membrane Protein G Family
		0017	(CT871)	pmp_4	PMP_4 (frame-shift with 0016)
	50	0018	(CT874)	pmp_5	Polymorphic Outer Membrane Protein G Family
		0019	(CT871)	pmp_5	PMP 5 (frame-shift with 0018)
		0444	(CT871)	pmp_6	Polymorphic Outer Membrane Protein G/I Family
		0445	(CT871)	pmp_7	Polymorphic Outer Membrane Protein G/F Parnity
		0446		· · -	
	55		(CT871)	pmp_8	Polymorphic Outer Membrane Protein G Family
	55	0447	(CT871)	pmp_9	Polymorphic Outer Membrane Protein G/I Family
		0450	(CT871)	pmp_10	Polymorphic Outer Membrane Protein G Family
		0449	(CT871)	pmp_10	PMP_10 (Frame-shift with 0450)

0636 (CT517) rl24

L24 Ribosomal Protein

0451

(CT871)

```
(CT071)
                                              CT071 Hypothetical Protein
                  0345
                                    yaeM
                                              YaeS family Hypothetical Protein
                        (CT450)
                                    yaeS
                  0566
                  0591
                         (CT472)
                                    yagE
                                              YagE family
                                              YbaB family Hypothetical Protein
                  0039
                        (CT335)
                                    ybaB
       5
                                    ybbP
                                              YbbP family Hypothetical Protein
                        (CT012)
                  0101
                                              10jap Superfamily Ortholog
                  0915
                        (CT769)
                                    ybeB
                                              ACR family
                  0137
                        (CT108)
                                    ybgI
                                              ATPase
                  0529
                        (CT402)
                                    ycaH
                                              PP-loop Superfamily ATPase
                  0438
                        (CT287)
                                    ycbF
      10
                  0734
                        (CT627)
                                    yceA
                                              YceA Hypothetical Protein
                                              Predicted Kinase
                  0954
                        (CT804)
                                    ychB
                                              PP-Loop Superfamily ATPase
                  0261
                        (CT217)
                                    ydaO
                        (CT127)
                                    ydhO
                                              Polysacchande Hydrolase-Invasin Repeat Family
                  0245
                                    yebC
                                              YebC Family Hypothetical Protein
                  0573
                        (CT457)
      15
                  0689
                         (CT687)
                                    yfhO_1
                                              NifS-related Aminotransferase 1
                  0862
                        (CT721)
                                    yfhO_2
                                              NifS-related Ammotransferase_2
                  0547
                         (CT434)
                                    ygbB
                                              YgbB Family Hypothetical Protein
                                              YggF Family Hypothetical Protein
                  0237
                         (CT184)
                                    yggF
                  0775
                         (CT606)
                                              YggV Family Hypothetical Protein
                                    yggV
      20
                  0396
                         (CT258)
                                    yhfO 3
                                              NifS-related Aminotransferase_3
yhhF
                                              Predicted Methylase
                         (CT487)
0605
                  0575
                         (CT458)
                                    yhhY
                                              Amino Group Acetyl Transferase
                         (CT473)
                                              YidD Family
                  0592
                                    yıdD
                  0982
                         (CT825)
                                    yıgN
                                              YigN Family Hypothetical Protein
      25
                                              YjeE Hypothetical Protein
                  0657
                         (CT537)
                                    ујеЕ
ļ.
                  0768
                         (CT644)
                                              Yohl Predicted Oxidoreductase
                                    yohI
M
                  0336
                         (CT077)
                                              YolL Hypothetical Protein
                                    yojL
W
                  0217
                         (CT140)
                                    ypdP
                                              YpdP Hypothetical Protein
14
                  0140
                         (CT212)
                                    yqdE
                                              YqdE Hypothetical Protein
ļ.
      30
                  0263
                         (CT221)
                                    yqfU
                                               YqfU Hypothetical Protein
į eis
                  0139
                         (CT211)
                                    yqgE
                                               YqgE Hypothetical Protein
j ub
                         (CT137)
                                    ywiC
                                              SuA5 Superfamily-related Protein
                  0270
ļ.
                  0879
                         (CT738)
                                    yycJ
                                              Metal Dependent Hydrolase
.II
T.
      35
                                                            Homologs to CHLTR Hypothetical Coding Genes
                  1000
                         (CT001)
                                    CT001 Hypothetical Protein
                  0020
                         (CT351)
                                    CT351 Hypothetical Protein
                  0021
                         (CT350)
                                    CT350 Hypothetical Protein
                                    CT345 Hypothetical Protein
                  0026
                         (CT345)
      40
                         (CT339)
                  0035
                                    CT339 Hypothetical Protein
                  0036
                         (CT338)
                                    CT338 Hypothetical Protein
                  0055
                         (CT296)
                                    CT296 Hypothetical Protein
                  0062
                         (CT289)
                                    CT289 Hypothetical Protein
                  0065
                         (CT288)
                                     CT288 Hypothetical Protein
      45
                  0068
                         (CT360)
                                    CT360 Hypothetical Protein
                  0071
                         (CT325)
                                    CT325 Hypothetical Protein
                         (CT324)
                  0072
                                     CT324 Hypothetical Protein
                  0085
                         (CT311)
                                     CT311 Hypothetical Protein
                  0087
                         (CT309)
                                     CT309 Hypothetical Protein
      50
                         (CT303)
                  0093
                                     CT303 Hypothetical Protein
                  0100
                         (CT011)
                                    CT011 Hypothetical Protein
                  0104
                         (CT017)
                                    CT017 Hypothetical Protein
                                     CT016 Hypothetical Protein
                  0105
                         (CT016)
                         (CT058)
                                     CT058 Hypothetical Protein_1
                  0107
       55
                         (CT018)
                                     CT018 Similanty
                  0108
                  0111
                         (CT021)
                                     CT021 Hypothetical Protein
                  0121
                         (CT031)
                                     CT031 Hypothetical Protein
```

		0441	(CT007)	CT007 Hypothetical Protein
		0442	(CT006)	CT006 Hypothetical Protein
		0443	(CT005)	CT005 Hypothetical Protein
		0474	(CT365)	CT365 Hypothetical Protein
	. 5	0476	(CT865)	CT865 Hypothetical Protein
		0480	(CT383)	CT383 Hypothetical Protein
		0485	(CT382)	CT382.1 Hypothetical Protein
		0487	(CT384)	CT384 Hypothetical Protein
			(CT386)	CT386 Hypothetical Protein
	10	0489		CT387 Hypothenical Protein
	10	0490	(CT387)	CT389 Hypothencal Protein
		0491	(CT389)	• •
		0496	(CT391)	CT391 Hypothetical Protein
		0497	(CT388)	CT388 Hypothetical Protein
	15	0506	(CT421)	CT421 Hypothetical Protein
	13	0507	(CT421)	CT421 1 Hypothetical Protein
		0508	(CT421)	CT421 2 Hypothetical Protein
		0512	(CT425)	CT425 Hypothetical Protein
		0514	(CT427)	CT427 Hypothetical Protein
344	20	0518	(CT429)	CT429 Hypothetical Protein
	20	0522	(CT433)	CT433 Hypothetical Protein
HI.		0525	(CT398)	CT398 Hypothetical Protein
254 254 254		0533	(CT406)	CT406 Hypothenical Protein
W		0537	(CT814)	CT814 1 Hypothetical Protein
Ü	25	0538	(CT814)	CT814 Hypothetical Protein
Į.i.	25	0554	(CT440)	CT440 Hypothetical Protein
		0559	(CT441)	CT441.1 Hypothetical Protein
i Fi		0565	(CT449)	CT449 Hypothetical Protein
B		0572	(CT456)	CT456 Hypothetical Protein
1-3-	30	0582	(CT465)	CT465 Hypothetical Protein
	30	0583	(CT466)	CT466 Hypothetical Protein
-L		0588	(CT469)	CT469 Hypothetical Protein
14		0589	(CT470)	CT470 Hypothetical Protein
		0590	(CT471)	CT471 Hypothetical Protein
T	35	0593	(CT474)	CT474 Hypothetical Protein
*******	33	0595	(CT476)	CT476 Hypothetical Protein
		0601	(CT483)	CT483 Hypothetical Protein
		0602	(CT484)	CT484 Hypothetical Protein
		0606	(CT488)	CT488 Hypothetical Protein CT490 Hypothetical Protein
	40	0609 0622	(CT490) (CT503)	CT503 Hypothetical Protein
	40	0623	(CT504)	CT504 Hypothetical Protein
		0648	(CT529)	CT529 Hypothetical Protein
	•	0658	(CT538)	CT538 Hypothetical Protein
		0667	(CT546)	CT546 Hypothetical Protein
	45			CT547 Hypothetical Protein
	43	0668	(CT547)	CT548 Hypothetical Protein
		0669	(CT548)	
		0671	(CT550)	CT550 Hypothetical Protein
		0673	(CT552)	CT552 Hypothetical Protein
	50	0675	(CT696)	CT696 Hypothetical Protein
	30	0676	(CT695)	CT695 Similarity
		0681	(CT691)	CT691 Hypothetical Protein
		0687	(CT482)	CT482 Hypothetical Protein
		0688	(CT481)	CT481 Hypothetical Protein
	<i>E E</i>	0700	(CT676)	CT676 Hypothetical Protein
	55	0705	(CT671)	CT671 Hypothetical Protein
		0706	(CT670)	CT670 Hypothetical Protein
		0708	(CT668)	CT668 Hypothetical Protein

		0709	(CT667)	CT667 Hypothetical Protein
		0710	(CT666)	CT666 Hypothetical Protein
		0711	(CT665)	CT665 Hypothetical Protein
		0713	(CT663)	CT663 Hypothetical Protein
	5	0717	(CT656)	CT656 Hypothetical Protein
		0718	(CT657)	CT657 Hypothetical Protein
		0720	(CT659)	CT659 Hypothetical Protein
		0722	(CT654)	CT654 Hypothetical Protein
		0725	(CT652)	CT652.1 Hypothetical Protein
	10	0726	(CT620)	CT620 Hypothetical Protein
		0727	(CT619)	CT619 Hypothetical Protein
		0739	(CT638)	CT368 Hypothetical Protein
		0742	(CT635)	CT635 Hypothetical Protein
		0746	(CT632)	CT632 Hypothetical Protein
	15	0747	(CT631)	CT631 Hypothetical Protein
		0751	(CT651)	CT651 Hypothetical Protein
		0755	(CT616)	CT616 Hypothetical Protein
		0760	(CT611)	CT611 Hypothetical Protein
		0761	(CT610)	CT610 Hypothetical Protein
113	20	0764	(CT648)	CT648 Hypotheti al Protein
		0765	(CT647)	CT647 Hypotheti al Protein
The little was straightful		0766	(CT646)	CT646 Hypothetical Protein
e e même		0767	(CT645)	CT645 Hypotheti al Protein
LLI aux		0770	(CT642)	CT642 Hypotheti al Protein
W	25	0774	(CT606)	CT606 1 Hypothetical Protein
		0776	(CT605)	CT605 Hypothetical Protein
m		0779	(CT602)	CT602 Hypothetical Protein
		0783	(CT592)	CT598 Hypothetical Protein
8		0791	(CT590)	CT590 Hypothetical Protein
Ē÷	30	0792	(CT589)	CT589 Hypothetical Protein
L		0803	(CT584)	CT584 Hypothetical Protein
		0807	(CT580)	CT580 Hypothetical Protein
ļ.b		0808	(CT579)	CT579 Hypothetical Protein
		0809	(CT578)	CT578 Hypothetical Protein
.II	35	0810	(CT577)	CT577 Hypothetical Protein
19-0-		0814	(CT573)	CT573 Hypothetical Protein
		0818	(CT569)	CT569 Hypothetical Protein
		0819	(CT568)	CT568 Hypothencal Protein
		0820	(CT567)	CT567 Hypothetical Protein
	40	0821	(CT566)	CT566 Hypothetical Protein
	. •	0822	(CT565)	CT565 Hypothencal Protein
		0827	(CT560)	CT560 Hypothetical Protein
		0834	(CT556)	CT556 Hypothetical Protein
		0840	(CT700)	CT700 Hypothetical Protein
	45	0842	(CT702)	CT702 Hypothetical Protein
	.5	0843	(CT702)	CT702 Hypothetical Protein
		0852	(CT711)	CT711 Hypothetical Protein
		0853	(CT712)	CT712 Hypothetical Protein
		0857	(CT716)	CT716 Hypothetical Protein
	50			
	50	0859	(CT718)	CT718 Hypothetical Protein
		0865	(CT724)	CT724 Hypothetical Protein
		0869	(CT728)	CT728 Hypothetical Protein
		0874	(CT733)	CT733 Hypothetical Protein
	55	0875	(CT734)	CT734 Hypothetical Protein
	55	0884	(CT741)	CT741 Hypothetical Protein
		0887	(CT744)	CHLTR Possible Phosphoprotei
		0896	(CT753)	CT753 Hypothetical Protein

	0906	(CT763)	CT763 Hy	pothetical Protein
	0908	(CT764)	CT764 Hy	pothetical Protein
	0912	(CT768)	CT768 Hy	pothetical Protein
	0925	(CT779)	CT779 Hy	pothetical Protein
5	0938	(CT788)	CT788 Hy	pothetical Protein
	0939	(CT790)	CT790 Hv	pothetical Protein
	0943	(CT794)	-	Typothetical Protein
	0945	(CT795)		pothetical Protein
	0956	(CT805)		pothetical Protein
10				
10	0960	(CT809)		pothetical Protein
	0989	(CT832)		pothencal Protein
	0994	(CT837)		pothetical Protein
	0995	(CT838)		pothetical Protein
1 6	0996	(CT839)		pothetical Protein
15	1002	(CT845)	CT845 Hy	pothetical Protein
	1003	(CT846)	CT846 Hy	pothetical Protein
	1004	(CT847)	CT847 Hy	pothetical Protein
	1005	(CT848)	CT848 Hy	pothetical Protein
	1006	(CT849)	CT849 Hy	pothetical Protein
20	1007	(CT849)	CT849 1 F	Typothetical Protein
	1008	(CT850)	CT850 Hy	pothetical Protein
	1010	(CT852)	CT852 Hy	pothetical Protein
	1011	(CT853)	CT853 Hy	pothetical Protein
	1015	(CT857)	CT857 Hy	pothetical Protein
25	1016	(CT858)	CT858 Hy	pothetical Protein
	1019	(CT860)	CT860 Hy	pothetical Protein
	1020	(CT861)	CT861 Hy	pothetical Protein
	1022	(CT863)	СТ863 Ну	pothetical Protein
	1032	(CT373)	CT373 Hy	pothetical Protein
30	1033	(CT372)	CT372 Hy	pothetical Protein
	1034	(CT371)	CT371 Hy	pothetical Protein
	1057	(CT356)	CT356 Hy	pothetical Protein
	1058	(CT355)	CT355 Hy	pothetical Protein
	1061	(CT330)	СТ330 Ну	pothetical Protein
35	1073	(CT371)	CT371 Hy	pothetical Protein
				Coding Genes Not in C. trachomatis
	0486			Hypothetical Proline Permease
	0279			Possible ABC Transporter Permease Protein
40	0505			3-Methyladenine DNA Glycosylase
	0193		argR	Similarity to Arginine Repressor
	1041		bioA	$Adenosylmethionine 8-Amino-7-Oxonon anoate\ Aminotransfer as e$
	1044		bioB	Biotin Synthase
	1042		bioD	Dethiobiotin synthetase
45	0585			Similarity to Cps IncA_2
	0562			CHLPS 43 kDa Protein Homolog_l
	0927			CHLPS 43 kDa Protein Homolog_2
	0928			CHLPS 43 kDa Protein Homolog_3
	0929			CHLPS 43 kDa Protein Homolog_4
50	1045			Conserved Hypothetical Membrane Protein
	0251			Conserved Hypothetical Protein
	0278			Conserved Outer Membrane Lipoprotein Protein
	0907			CutA-like Periplasmic Divalent Cation Tolerance Protein
	0171		guaA	GMP Synthase
55	0172		guaB	Inosine 5'-Monophosphase Dehydrogenase
•	0608		-	Undine 5'-Monophosphate Synthase
	0735			Undine Kinase
				- 1

0906 (CT763) CT763 Hypothetical Protein

Here and the maje that when the maje

		0980		Similar to Sacc	haromyces cerevisiae 52 9KDa Protein	
		0232		Similarity to 5'-Methylthioadenosine Nucleosidase		
		1046		Tryptophan Hy	droxylase	
		0477	ygeV Bs	Conserved Hyp	pothetical Protein	
	5	0048	yqfF-Bs		pothetical IM Protein	
		0587			pothetical Protein	
		0143	yxjG_Bs_		Hypothetical Protein	
		0448	yxjG_Bs_		Hypothetical Protein	
		0006	0180	0440	0977	
	10		0181	0455	0978	
	10	0007	0190 .	0456	1018	
		0008	0203	0457	1023	
				0458	1027	
		0010	0204	0459	1029	
	15	0011	0205			
	15	0012	0209	0460	1040	
		0028	0210	0461	1051	
		0029	0211	0462	1052	
		0034	0212	0463	1053	
	20	0041	0213	0464	1054	
13	20	0042	0214	0465	1055	
157		0043	0215	0472	1056	
The state of the s		0044	0216	0473	1064	
		0045	0218	0481	1065	
m	2.5	0046	0220	0483	1066	
E.S.	25	0047	0221	0492	1070	
m	30	0049	0222	0493	1071	
		0050	0223	0494	1072	
		0051	0224	0498		
e La		0063	0225	0499		
14		0064	0226	0516		
		0066	0233	0517		
		0067	0240	0523		
		0069	0241	0524		
H	0.5	0070	0242	0553		
	35	0099	0243	0574		
		0124	0266	0600		
		0125	0267	0656		
		0126	0268	0664		
	40	0130	0277	0677		
	40	0131	0283	0678		
		0132	0284	0685		
		0142	0285	0686		
		0146	0287	0724		
		0147	0352	0731		
	45	0155	0353	0745		
		0156	0354	0753		
		0157	0355	0794		
		0158	0356	0795		
		0159	0357	0796		
	50	0162	0358	0797		
		0163	0365	0798		
		0164	0366	0799		
		0165	0367	0829		
		0166	0368	0830		
	55	0167	0371	0831		
		0168	0372	1880		
		0169	0375	0882		

	0170	0376	0913
	0173	0391	0914
	0174	0398	0930
	0175	0404	0944
5	0177	0431	0964
	0178	0432	0975
	0179	0439	0976

Chlamydia pneumoniae Genome Encoded Proteins

CPn_0001 330 4
CT001 hypothetical protein
KRLKDEIKYTSLRRKAMLGKIIRGLSSLIVILCALNVGLIGITHNKLNIIAKLCGGVSTP ATQITYIIIGIAGVICLLSFCPFCSKKSRHSHGDSCSSGGCHSHHSDKN

570 CPD_000: 579 ACC ACC TO BE ACC TO BE ACC TO BE ACCOUNTED TO BE ACCOUNTED.

889

CPn_0003 889 2370
gata-Glu trna Gln Amidotransferae
kIMYRYSALELAKAVTUGELTATGYTQHFFHRIEAEGQVGAFISLCKEQALEQAELIDK
KRSRGEPLGKLAGVPVGIKDNIHVTGLKTTCASRVLENYQPPFDATVVERIKKEDGIILG
KLNMDEFAMGSTTLYSAFHPHNPMDLSRVPGGSSGGSAAAVSARFCPVALGSDTVGGSIR
QPAAFCGVVGFKPSYGAVSRYGLVAFASSLDQIGPLANTVEDVALMMDVFSGRDPKDATS
REFFRDSFMSKLSTEVPKVIGVPRTFLBGLRDDIRENFFSSLAIFEGESTHLVDVELDIL
SHAVSIYYILASAEAATNLAFPGVRYGYRSPQAHTISQLYDLSRGEGFGKEVMRILLG
NYVLSAERQNVYYKKATAVRAKIVKAFRTAFEKCEILAMPVCSSPAFEIGEILDPVTLYL
QDIYTVANNLAYLPALAVPSGFSKEGLPLGLQIIGQQGQDQQVCQVGYSFQEHAQIKQLF
SKRYAKSVVIGGOS SKRYAKSVVLGGQS

CPn_0004 2334 3833
gatB-(Pet112) Glu tRNA Gln Amidotransferase (B Subunit)
EICQKCCSRRSIMSAVYADWESVIGLEVHVELNTASKLFSSALNRFGDEPNTNISTVCTG
LPGSLPVLNQSAVEKAVLFGCAVEGEISLLSRFDRKSYFYPDSPRNFQITQFEHPIIRGG
RIKAIVQGERYYFELAQTHIEDDAGMLKHFGEFAGVDYNNAGVPLIEIUSKPCMFCPEDA
VAYATSLVSLLDYIGISDCNMEEGSIRFDVNVSVRPKGSPELRNKVEIKNMSFAFMAQA
LEAEKQRQIDEYLNQPNKDPKLVIFAATYRWDPEKKKTVLMRLKESAEDYKKYFPEPDLPT
LQLTESYIERIRKTLPELPYDKYHRYIQEYGLSEDIASILISDKNIATFFEVACKDCKNF
RSLSMWVTVEFGGRCKTIGVKLPSSGIFFEGVAQLVNAIDQGVITGKIAKEIADLMMESF
GKNPEEILKEKPELLPMSDEGELQKIIAEVVLANPESIVDYKNGKTKALGFLVGQIMKRT
AGKAPPKRVNELLLLELDKG

CPn_0005 4097 6892 CPT_0005 4097 6892

pmp_1-Polymorphic Outer Membrane Protein

SDIHFDLGTKMRFSLCGFPLVFSPTLLSVFDTSLSATTISLTPEDSFHGDSQNAERSYNV

QAGDVYSLTGDVSISNVDNSALNKACFNVTSGSVTFAGNHHGLYFNNISSGTTKEGAVLC

CQDRQATARFSGFSTLSFIQSFGDIKEQGCLYSKNALMLLNNYVVRFEQNGSKTKGGAIS

GANATIVGNYDSVSFYQNAATFGGAIHSSGPLQIAVNQAEIRFAQNTAKNGSGGALYSDG

DIDLEQNAVVLFRENEALITAIGKGGAVCCLPTSGSSTPVBIVTFSDNXQLVFERNHSIM

GGGATYARKLSISSGGFTLFINNISYANSQNLGGAIAIDTGGEISLSAEKGTITFQGNRT GGSÄYARKLSISSGGPTLFINNISYANSONLGGAIAIDTGGEISLSAEKGTITFQGNRT
SLEFINGIHLLQNAKFLKLQARNGYSIEFYDPITSEADGSTOLNINGDPKNKEYTGTILF
SGEKSLANDPRDFKSTIPQNVNLSAGYLVIKEGAEVTVSKFTQSFGSHLVLDLGTKLIAS
KEDIAITGLAIDIDSLSSSSTAAVIKANTANKQISVTDSIELISPTGNAYEDLRMRNSQT
PPLÖSLEPGAGGSVTVTVAGDFLPVSHYYGFQGMWKLANTGNKVGEFFWDKINYKPRPE
KEGNLVPNILWGNAVDVRSLMQVQETHASSLQTDRGLWIDGIGNFFHVSASEDNIRYRHN
SGĞVLSVNNEITPKHYTSMAFSQLFSDKDVAVSNNEYRMYLGSYLYQYTTSLGNIFRY
AŞRMPNVNYGLISRRFLONPLMIFHFLCAYGHATNDMKTDYANFFMVKNSWRNNCWAIEC
GGSMPLLVFENGRLFQGAIPFMKLQLVYAYQGDFKETTADGRRFSNGSLTSISVPLGIRF
EKLÄLSQDVLVPSFSYIPDIFRKDPSCEAALVISGDSWLVPAAHVSRHAFVGSGTGRYH
FNDFYEELLCRGSIECRRPHARNYNNGGSKFRF FNDYTELLCRGSIECRPHARNYNINCGSKFRF

CPn_0006 7299 7141
No robust homolog present in Genebank/EMBL as of 11/7/98
KQLQEPLRSALLERLSEWLVLLGVPSPETTRSTPEKDANQLPKDSRNRTLESL

7488 10496 No robust homolog present in Genebank/EMBL as of 11/7/98 KSEAYNLSLIFSFLVVIPLTDSTTSSLSTSLLDEGNPQSMRKLRILAIVLIALSIILIAG KSERMNLSLIFSFLVVIPLTDSTTSSLSTSLLDEGNPOSMRKLRILAIVLIALSIILIAG
GVYLLTVAIFGLSSVISSPAGMGACALGCVMLALGIDVLLKKREVPIVLASVTTTPOTGS
FRSGESISGADSTIRSLPTYLLDEGHPOSMRKLRILAIVLIVFSIILIASGVVLITVAIP
GLSSVISSPAGMGACALGCVMLALGIDVLLKKREVPIVLASVTTTPOTGSPRSGISISGA
DSTERSLPTYPLDEGHPOSMRKLRILAIVLIVFSIILIASGVVLLTVAIPGLSSIISSPA
EMGACALGCVMLALGIDVLLKKREVPIVVPAPIPEEVVIDDIDEESIRLOQEARAALARL
PEEMSAFEGYIKVVESHLENMKSLPYDGHGLEEKTKHQIRVVRSSLKAMVPEFLDIRRIF
EEEEFFFLSARKKLIDLATTLVERRILTEGLERNNLRKAFSYLYQDSIFKKIIDNFEKLA
WKFMILSKSICRFTIIFENHEHGVAKSLLHKNAVLLEKVIYRSLQKSYRDIGMSSAKMKI
LHGNPFFSLEDNKKTINKEHAEMLESLSSYRKVFLALSDENVVDTPSDPKKNDLSGIPCR
DALSEISRDEQWGKKAHLKHQESLYTQAADRITDGSSKENOKELEKAEQEYISSWERVK
FEIERVQERIRAIQKLYPNILEREETTGQETVTPTVOGTTASSOLTDILGRIEVSSRED
NONQESCVVLASHEVEMSWEVKQEYGPKKKEFQDQMGSLERFFTEHIEELEVLOKDYSK
HLSYFKKVNNKKEVQYAKFRLKVLESDLEGILAQTESAESLLTQEELEILATRGALEKAV
FKGSLCCALASKAKPYFEEDPFFQDSTOURALTLRLQEAKASLEEEIKRFSNLENDIAE
ERRLLKESKGYTERRAGIGVLREIAVESTYDLASLITNTWECTPESEKVYFSMYLNYYNEEK
RRAKTRLVEMTORYRFKMALEAMQFNEEALLQEELSIQAPSE

CPn_0008 10780 11685 No robust homolog present in Genebank/EMBL as of 11/7/98 CKYSYLLMYPPPPRRSLOVSCSKLRSLSITLLVLGVLLLTLGIPGLTAGISFGAGLGFSA LGGVLVISGLLFLLVRREVPTVRSEEIPRGVSVTPSEEPALEKAGKEPETKKILDRLPKE LDQLDTY1QEVFACLERLKDPKYEDRGLLTEAKEKLRVFDVVEKDMYSEFLD1QRVLMEE AYYVEHCQDPLEN1AYE1FSSQELRDYYCAGVCGYLPSGDARADRLKRSVKEVMDRFMRV TWRSWEASVMLDHSYGVARELFKKAVGVLEESVYK1LFKSYRDAFYECEKAK1QRDGRFK

RRAKTRLVEMTQRYRDFKMALEAMQFNEEALLQEELSIQAPSE

CPn_0009 L1689 L3119
No cobust homolog present in Genebank/EMBL as of 11/7/98
I/JSAIIAEORERDINGUMEDLROTIEWVGEHDCTDIETVRKJCMWIDRYADKFILREKEEK
MERHELFHATMVRKASGHAYAKAKAAFEKERSNENOPKVKDVEKWLSKGLAEFRNOESRR
ARERLÆLOTLYPEVSVEERVLERORTKKVNLENLYADIEKKYHHKVREQEHYMKEVENK
FAETKENGEKVLJAEEVJECLORLEDGLETWUKKLTKAEEJVFEMKFDATEKLGNKVLSD
VINRLEILJEDALEMIFRIEEIEMTLRHVELPLLEMKNTFEKASLQYNSCKEMLAKVEPO
CKESFTYRHJORRLERLNODLGTAYTNOCOBRIGGTODLEGKWTCROHLREDMKHFEVOG
LNFINEELLWVJAELFDGARLDLVATVPYMEPYLQYHNIKREKVRHQWMAKTERYREIRQ
AFQXVMKEDLLAEDTILKEEDIWLLRDDWLLRDERKNRORRLICNKIAAAOOPVKGF

CEn_0010 13324 14325 No robust homolog present in Genebank/EMBL as of 11/7/98 CKYFYLRSYPPPPOHSVGSIGSPSKLRVLAITFLVFGMLLLIGGALFLTLGIPGLSAAIS FGLDIGLSALGGVLMISGLLCLUVKREIFTVRPEEIFEGVSLAFSEERALGAAQKTLAQL PKELDQLDTDIGEVFACLRKLKDSKYESRSFLNDAKKELRVFDFV/EDTLGEIFELRQIV AQEOMOLDHILINGGRSLMMTAESSELDLFHVSKRLGYLPSGDVRGEGLKKSAKEIVARLM SLHCEIHKVAVAFDRNSYAMAEKAFAKALGALEESVYRSLTQSYRDKFLESERAKIPWNG HITWLRDDAKSGCAEKKLGMPRNVGRNLGKQSFG

CPn_0010 1 14268 15746

TLABEL DESTRUCTION OF THE STANDARD OF PEWINDERSON TRENKIQET RSNLEKAYEA I EENYRCCVREQEDYWKEEEKREAEFREROKKIL
SPEELESSLEQFDHGLKNFSEKLMELEGH ILKLQKEATAEVENKILSDAESRLEIVFEDV
KEMPCRIEEIEKTLRMAELPLLPTKKAFEKACSOVINSCAEMLEKVKPYCKESLAYVTSKE
RLVSLIEDLDRRAYTEGOKRPGGDSGLESEVRACREQLERERIQEFETGGLDLVEKELLCVS
SRLNNTECDCVSGVKKEAPPGKKFYAQYYDEIYRVRVQSRWMTMSERLREGVQACNKMLK
AGLSEEDKVLKEEEYWLYREERKNKEKRLVGTKIVATQQRIQEFQPSDIVESSNEKVSLM DKARFLFNREDHS

CPn_0011 15877 16614
gatB-(Pet112) Glu tRNA Gln Amidotransferase (B Subunit)
fWYSIMTAAPAILHVSPTPPEETKFVIPKDSKSRALGITLLVVGILLVVCGAIVLSGVIS
GLSALIVCGLGISTISLGVVLFYLGILLLLKREITLEQ LEAKQIAETFADELKELEMYI
QSTEKSLEKIEDSRYSDQGFLNRATQKILDLESSLSSITSEFRDLRQLFDEEKIELLSGE
RLLEFIAANLFKQGRDVYLNLGNLADIRAYMGPNNYKVAMVIEKAKAVVHEFIVLTTMAR
FLEFEF

CPn_0012 16596 18212
gatB-(Pet112) Glu tRNA Gln Amidotransferase (B Subunit)
GIRVFFLKNKYGLLKGMYQENLRLLERLLYNSVQKSYADRLFSVEKTKNYHDTPLIFWEE
DKEKCAEAEKAFLEQOKILLDYGKSIFWLNENDEINLINDFWSWGLNTVRTRKVFQEVDDS
ERWNHKVLIQKLEDDVEKLLEESSKESTEANKKLLSDLVDRLEDAKTKFFLKKQEEVETR
VKDLRARYGGTVDPKQDTEAKKKVELEASLETFLDSIESELVQCLEDQDIYWKEQDVKDL
ARTOELEEQDIEAKREEAAEDLRSLNERLKKSKTMLDRAKWHIENAEDSITWMTSQIEMK
DMKARLKILKEDITSVLPEIDEI ETCLSLEELPLLTTRELLTKSYLKFKIGSETLLKMTS
VFENNIYVQEYEVQLQNLGFKLQGISQRFGKKQDDFANLEEQVALQKKRLRELTQNFEIQ
GFNFMKEDFKAAAKDLYIRSTAEQKMMFDVPCMELFRRYHEEVNKPLLELMYNCADSYRD AKKKLCSLRLDEKELLQKEIKKEEFYQKKQQRHADRSRHTTYQKLRIAEELALELKKKI

CPn_0013 18509 21106
pmp_2-Polymorphic Outer Membrane Protein
LRDRLAFFIYLLYWKESPLREKKVVMKIPLRFLLISLVPTLSMSNLLGAATTEELSASNS
FÖGTTSTISTSSKKTSSAIDGINYVFKDSVVIENVPKTGETQSTSCFKNDAAAGDLNFLGG
GFSFTFSNIDATTASGAAIGSEAANKTVTLSGFSALSFLKSPASTVTNGLGAINVKGNLS GFSFTFSHIDATIASGAAIGSEAANKIVTLSGFSALSFLRSFASTVIRGLGAINVRUNLS
LLDNDKVLIQDNFSTGDGGAINCAGSLKIANNKSLSFIGNSSTRGGAIHTKNILTLSSG
ETLFQGNTAPTAAGKGGAIAIADSGTLSISGDSGDIIFEGNTIGATGTVSHSAIDLGTSA
KITALRAAQGHTIYFYDPITVTGSTSVADALNINSPDTGDNKEYTGTIVFSGEKLTEAEA
KDEKNRTSKLLQNVAFKNGTVVLKGDVVLSANGFSQDANSKLIMDLGTSLVANTESIELT
NLEINIDSLRNGKKIKLSAATAQKDIRIDRFVVLAISDESFYQNGFLMEDHSYDGILELD
AGKDIVISADSRSIDAVQSPYGYQGKWTINWSTDDKKATVSWAKQSFMPTAEQEAPLVPN LLWGSF IDVRSFONF IELGTEGAPYEKFFWVAGISNVLHRSGRENQRKFRHVSGGAVVGA STRMFGGDTLSLGFAQLFARDKDVFMNTNFAKTYAGSLRLQHDASLYSVVS ILLGEGGLR EILLPYVSKTLPCSFYGQLSYGHTDHRMKTESLPPPPPTLSTDHTSWGGYVAAGELGTRV AVENTSGRGFFQEYTPFVKVQAVYARQDSFVELGAISRDFSDSHLYNLAIPLGIKLEKRF AEQYYHVVAMYSPDVCRSNPKCTTTLLSNQGSWKTKGSNLARQAGIVQASGFRSLGAAAE LFGNFGFEWRGSSRSYNVDAGSKIKF

CPn_0014 21365 21922
pmp_3-Polymorphic Outer Membrane Protein
IQNOSIYFTMKSSFPKFVFSTFAIFPLSMIATETVLDSSASFDGNKNGNFSVRESQEDAG
TTYLFKGNVTLENIPOTGTAITKSCFNNTKGDLTFTGNGNSLLFGTVDAGTVAGAAVNSS
VVDKSTTFIGFSSLSFIASPGSSITTGKGAVSCSTGSLSLTKMSVCSSAKTFQRIMAVLS

CPn_0015 21835 24174
pmp_3-PMP_3 (frame-shift with 0014)
LEFDKNVSLLFSKNFSTDNGGAITAKTLSLTGTTMSALFSENTSSKKGGAIQTSDALTIT
GNGGVSFSDNTSSDSGAAIFTEASVTISNNAKVSFIDNKVTGASSSTTGDNSGGAICAY
KTSTDTKVTLTGNQMLLFSNNTSTTAGGAIYVKKLELASGGLTLFSRNSVNGGTAPKGGA
IAIEDSGELSLSADSGDIVFLGNTVTSTTPGTNRSSIDLGTSAMMTALRSAAGRAIYFYD
PITTGSSTTVTDVLKVNETPADSALQYTGNIIFTGEKLSETEAADSKNLTSKLLQPVTLS
GGTLSLKHGVTLQTQAFTQQADSRLEMDVGTTLEPADTSTINNLVINISSIDCAKKAKIE
TKATSKNLTLSGTITLLDPTGTFYENHSLRNPQSYDILELKASGTVTSTAVTPDPIMGEK
FHYGYGGTGGFIVMGTGASTTATFNWTKTGY JTPNPERIGSLVPNSLWNAFIDISSLHYLM
ETANEGLQGDRAFWCAGLSNFFHKDSTKTRRGFRHLSGGYVIGGNLHTCSDKILSAAFCQ
LFGBRDVFVAKNQGTVYGGTLYYQHNETYISLPCKLRPCSLSVVPTEIPVLFSGNLSYT
HTDNDLKTKYTTYPTVKGSMGNDSFALEFGGRAPICLDESSALFEQYMPFMKLGFYVYANQE
GFKEQGTEAREGSSSLVNLALFIGIRFDKESDCQDATYNLTLGYTVDLVRSNPDCTTTL
RISGDSWKTFGTNLARQALVLRAGNHFCFNSNFEAFSQFSFELRGSSRNYNVDLGAKYQF

CPh_0016 24383 26188
pmp_4-Polymorphic Outer Membrane Protein
RSDFALKROCHRRSSFSLLLISSSLAFPLLMSVSADAADLTLGSRDSYNGDTSTTEFTPK
RATSDASGTTYILDGDVSISJAGKOTSLTTSCFSNTAGNLTFLGNSFSLHFDNIISSTVA
GVVVSNTAASGITKFSGFSTLRMLAAPPTTCKGAIKITDSLVFESIGNLDLNENASSENG
GAINTKTLSLTGSTRFVAFLJNSSSQQOGAIYASGDSVIGENAGILSFGNNSATTSGGAI
GAEGNLVISHNONLIFFDGCKATTINGAIDCHKAGAMPDPILTLESNDESHFLNNTAGNSG
GAIYTKKLVLSSGRGGVLFSNNKAANATPKGGAIAILDSGEIGISADLGNIIFEGNTTST
TGGFAGVTRNAIDLASNAKFLNLRATRGIRVIFFYDPITSSGATDKLSLNKADAGGGNTYE
GYTVPSGERLSEELEKKPDDNLKJTPTGAVELAGAGLVLKUJVTVV-NTITQVEGSKVVMD
GYTTTFAJARGVTLNSILAINIDGLDGTUPAILKATAGKDVALAGPINLVTAQGNYYEHH
HLGOQQVEPLIELSAQGTMTTTSIPDTPILPTTNIYGYQGSNNNNKTLRRPCNCKNKKCYLN
LD

POEKERS JEGSERO SEDTE UTVANS SEKES DEGLADOT TVANS DE KUDECCLIETE UTVANS DE KUDECCHARTEN AND TRUDECCLIETE UTVANS DE KUDECCHARTEN AND TRUDECCHARTEN AND TRUDEC

PGIEVFSHGAIELRGSSRNYNINLGGKYRF

CPn_0018 27513 29003
pmp_5-Polymorphic Outer Membrane Protein
EYMMKTSVSMLLALLCSGASSIVLHAATTPLNPEDGFIGEGNTNTFSPKSTTDAAGTTYS
LTGEVLYIDPGKGGSTTGTCFVETAGDLTFLGNGNTLKFLSVDAGANIAVAHVQGSKNLS
FTDFLSLVITESPKSAVTTGKGSLVSLGAVQLQDINTLVLTSNASVEDGGVIKGNSCLIQ FTDELSLYTESPRSAVTRANSLIDAVQUQUINTAVELISAVSUBDAV RANSELEAV OTKINSATEOQNISSKKOGA LOTTOOLUTENILOTUKENENKAVISGGALDIGASSISTAN HILLEVQNETOGIA NEMA (IF 1886 FITHET L. LIZEN FIMO COSALO FITTISI ITOSI DILIVE RITTOOLUTEN ALDANIE IL 1872 FILADE HILLIANDI VALLETINITOSI LOTET GGDIVFEGNOVITTAPNATIKRNVIHLESTAKWIGLAASQONAIYFYDEITINITOSISDN LRINEVSANQKISGSIVFSGERLSTAEATAENLISRINQFVILVEGSLVLKQGVILITQG FSQEPESTILLIDIGTSL

CPn_0019 29007 30356

pmp_5-PMP_5 (frame-shift with 0018)
ASTEDIVITNLSINADTIYGKNPINIVASAANKNITLTGTLALVNADGAFYENHTLQDSQ
DYSFVKLSPGAGGTIITQDASQKPLEVAPSRPHYGYQGHWNVQVIFGTGTQPSQANLEWV
RTGYLPNERQGSLVPNSLWGSFVDQRAIQEIMVNSQILCQERGVWGAGIANFLHRDKI
NEHGYRHSGVGYLVGVGTHAFSDATINAAFGCLFSRKDYVVXSNHGTSYSGVVFLEDTL
EFRSPQGFYTDSSSEACCNQVVTIDMQLSYSHRNNDMKTKYTTYPEAGSWANDVFGLEF
GATTYYYPNSTFLFDTYSPFLRLQCTYAHQEDFKETGGEVRHFTSGDLFNLAVPIGVKFE
RFSDCKRGSYELTLAYYPDVIRKDPKSTATLASGATWSTHGNNLSRQGLQLRLGNHCLIN
BOLEVESUGA IFLRGSSRNYNING GGKYRF

CPn_0020 32717 30603
Predicted OMP (leader (14) peptide: outer membrane)
KLWSNPNLRLMKRCFLFLASFVLMGSSADALTHQEAVKKKNSYLSHFKSVSGIVTIEDGV
LNTHNNLRIQANKVYVENTVGQSLKLVAHGNYWNYRAKTLVCDVLEYYEDTDSCLLTNG INIHNILRIQANKVYVENTVGQSLKLVAHGNVMVNYRAKTLVCDYLEYYEDTDSCLLITNG
RFAMYPWFLGGSMITLTPETIVIRKGYISTSEGPKKDLCLSGDYLEYSSDSLLSIGKTTL
RVCRIPILFLPPFSIMPMEIPKPPINFRGTGGFLGSYLGMSYSPISRKHFSSTFFLDSF
FKHGYGMGFNLHCSQKQVPENVFNMKSYYAHRLAIDMAEAHDRYRLHGDFCFTHKHVNFS
GSYHLSDSWETVADIFPNNFMLKNTGPTRVDCTWNDNYFEGYLTSSVKVNSFQNANQELP
YLTLRQYPISIYNTGYYLENIVECGYLNFAFSDHIVGENFSSLRLAARPKLHKTVPLPIG
TLSSTLGSSLIYYSDVPEISSRHSQLSAKLQLDYRFILHKSYIORRHIIEPFVFFITETR
FLAKNEDHYIFSIQDAFHSLNLLKAGIDTSVLSKTNPRFPRIHAKLWTTHILSNTESKPT
FPKTACELSLPFGKKNTVSLDAEWIWKHCWDHMNIRWEWIGNDNVANTLESLHRSKYSL
IKCDRENFILDVSRPIDQLLDSPLSDHRNLIIGKLFVRPHFCWYRLSLRYGWHRQDTPN
VLEYMBILGTXIEPEWAUD VGAVEDBEADGEBEREINIDERFER HENDENDERDER YLEYOMILGTKIFEHWQLYGVYERREADSRFFFFLKLDKPKKPPF

CFT 0021 34470 32707

Prédicted OMP [leader (19) peptide]

CSASPYPNIEILARGVEHRSMSLFHLTLFGLLLCSLPISLVAKFPESVGHKILYISTOST

QQALATYLEALDAYGDHDFFVLRKIGEDYLKQSIHSSDPQTRKSTIIGAGLAGSSEALDV

LSGAMETADPLQQLLVLSAVSGHLGKTSDDLLFKALASPYPVIRLEAAYRLANLKNTKVI

DH_HSFIHKLPEEIQCLSAAIFLRLETEESDAYIRDLLAAKKSAIRSATAQIGEYQQKR

FLEPELRIKLLTSASPQDQEAILYALGKLKDGQSYYNIKKQLQKPDVDVYLAAQALIALGK

EEDALPVIKKQALEERPRALYALRHLPSEIGIPIALPIFLKTKNSEAKLNVALALLELGC

DTPHCLEYITERLVQPHYNETLALSFSKGRTLQNWRRVNIIVPQDPQERERLLSTTRGLE

EGELFFLFREPKEAYLDCIYKLLASQKTQLATTAISFLSHTSHQEALDLLFQAAKLPGEP

ITRAYADLAIYNLTKDPEKKRSLHDYAKKLIQETLLFVDTENQRPHPSMPYLRYQVTPES

RIKKALDILETLATSKSSEDIRLLIQLMTEGDAKNFFVLAGLLIKIVE

CPn_0022 35042

TIŁQVISNCCNVSNTRSFYSMSLPLVLGSSSPRRKFILEKFRVPFTVIPSNFDESKVSYS GDPIAYTQELAAQKAYAVSELHSPCDCIILTGDTIVSYDGRIFTKPODKADAIQMLKTLR NQTHDVVTSIAVLHKGKLLTGSETSQISLTMIPDHRIESYIDTVGTLNNCGAYDVCHGGL ILKKVHGCVYNVQGLPIQTLKYLLEELNIDLWDYSI

CPb_0023 36657 35014 yjjK/alr-ABC Transporter Protein ATPase ENBAKLLYSKQHFVMLSAMSIVLDKIGKSLGTRILFDDVSVVFNPGNCYGLTGPNGAGKS ENRAKLLYSKOHFVMLSAMSIVLDKIGKSLGTRILFDDVSVVFNRGNCYGLTGPNGAGKS
TLLKIIMGMIEPTRGSISLPKKVGILRQNIDSFHDTTVLDCVIMGNTRUMALQRRDNLY
LOEFTDAIGMELGEIEEIIGEENGYRADSEAEELLTGIGIPNEMFDKKMAMIPIDLQFRV
LLCQALFGHPEALLLDEPTNHLDLYSINWLGNFLKDVEGTVIVVSHDRHFLHTITTHIAD
IDYDTIITYPGNYDDWYEMKTASROEKADIKSKEKKISOUKEFVAKFGAGSRASOVOSR
LREIKKLOPGELKKSNIQRPYIRFPLSDKSSGKVVLSLEAITKDYGDHQVIHPPSLEIYQ
GDKLGIIGNNGLGKTTLMKLLAGVEAPSSGSIKLGAICSYPONHSDVLADCGCETLF
EWLRNRKTGINDQEIRSVLGKMLFGGDDAFKQIQALSGGETARLLMAGMMLENHNVLILD
EANNHLDLESVSALSWAINDYKGTAIFVSHDRGLIQDCATKLLIFDKDKITFFDGTMVDY

37605 36661 xerC-Integrase/recombinase

XetC-Integrase/recombinase
REWMIASIYSFLDYLKMYKSASPHTLRNYCLDLNGLKIFLEERGNLAPSSPLQLATEKRK
VSELPFSLFTKEHVRMYIAKLIENGKAKRTIKRCLSSIKSFAHYCVIQKILLENPAETIH
GPRLPKELPSPMTYACVEVMATPDISKYHGLRDRCLMELPYSSGLRISEIVAVNKODFD
LSTHLIRIRGKKKERIIPYTSNAIQWIQIYLHYPDRKILEKDPQAIFLNRFGRRISTRS
LDRSFQEYLRRSGLSGHITPHTIRHTIATHWLESGMDLKTIQALLGHSSLETTTVYTQVS VKLKKOTHQEAHPHA

CPn_0025 38610 37584
elaC/atsA-Sulphohydrolase/Glycosulfatase
ILMSSRELIILGCSSQQPTRTRNQGAYLFRWNGEGLLFDPGEGTQROFIFANIAPTTVNR
IFVSHFHGDHCLGLGSMLMRLNLDKVSHFIHCYYPASGKKYPDRLRYGTIYHETIQVVEH
PISEEG IVEDFGSFRIEAQRLOHQVDTLGWRITEPDTIKFLPKELESRGIRGLIGDLIR
DQEISIGGSTVYLSDVSYVRKGDSIAIIADTLPCQAAIDLAKNSCMMLCESTILEQHRHL
AESHFHMTAKQAATLAKRAATQKLILTHFSARYLNLDDFYKEASAVFPNVCVAQEYRSYP EDKNDLLNK

CPn_0026 34637 38762 CT345 hypotherical protein CNFAMCHLIPSCRUCVTSYFHKPOPIKONAPSKJIRDICNINYLLICVLVVVVLVGAML

CMFTTERVETPULEDTALLEDGESTENDELINWTSTREPRETAPROADSOCPTSOABROS TQUSTBUILDIETRINFTRESKRIVETPTELIKKKRKTRIJELDLIRITSTETPTSOABROS ROBELLIPPARICKAPITYNTEKESKRIVETPTELIKKKRKTRIJELDLIRITTSOETHER SAPKPMYPSK KREVELIKKTYPLPTEDLEHQSISMESSDESTPPPVORKATLEWECKOPTDP

CPn_0027 42252 30778 **3977**\$\fonction ACD dependent Protected

PSIRTIVDSTTNSDSPILDFNFUPVEKLLDEGEEEGEDOSTERLLPSELFILPKNRPFF
PGMAAPILIESGPYYEVLKVLAKSSOKYIGLVLTKKENADILKVSFNQLHKTGVARILR
IMPIEGSSQVLLSIEERIRIIEFIKDKYLKARVSYHADNKELTEELKAYSISIVSVIKD
LLKLNPLFKEELQIFLGHSDFTEFOKLADFSVALTTATREELGEVLETTNMHDRIDKALI
LLKKELDLSRIOSSINGKIFATITKSGKEFFLKEQLKTIKKELGLEKEDRAIDIEKFSER
LKRRHVPDVAMEVIODEIEKLOTLETSSAEYTVCRNYLDMLTIIPMGIQSKEYHDLKKAE
IVLNKDHYGLDEIKQRILELISVGKLSKOLKGSIICLVGPPGVOKTSIGRSIAKVLHRKF
RFSVGGMRDEAEIKGHRPTYIGAMPGKMVOALKQSOAMNPVIMIDEVDKIGASYHGDPA

LUNTCONKOELEN PROFESSOKITTSTADALOTIPDFUSPMTIFELTOTLE
LAARIUMBARREN AUNT SAARIUMTKARATARTOTERDTUSPMTIFELTOTLE
LKIVONQEKPKSKKITFKISKNLOTYLGKPIFSSDRFYESTPVGVATGLAWTSLGGATL
VIESVOVSSLKTDMHLTGOAGEVMKESSQIAWTYLHSALHRYARGYTFFPKSQVHIHIPE YIESVOVSSLKTOMHLTOQAGEVMKESSÇIAWTYLHSALHRYAPGYTFFPKSOVHIHIPE GATPKDGPSAGITMVTSLLSLLLETPVVNNLGMTGEITLTGRVLGVGGIREKLIAARRSR LNILIFPEDNRRDYEELPAYLKTGLKIHFVSHYDDVLKVAFPKLK

43328 42543 CPH_0028 43328 42543
No robust homolog present in Genebank/EMBL as of 11/7/98
RMFLQFFHPIVFSDQSLSFLPYLGKSSGIIEKCSNIVEHYLHLGGDTSVIITGVSGATFL
SVDHALPISKSEKIIKILSYILILPILLALFIKIVLRIILFFKYRGLILDVKKEDLKKTL
TPDQENLSLPLPSPTILKKIHALHILVRSGKTYNELIQEGFSFFKITDLGQAESPKQDIG
FSYNSLLPNFYFHSLVSVPNISGEERALNYHKEQQEEMAVKLKTMQACSFVFRSLHLPSM QTKDKKAGFGLLTFFPWKIYPL

CPn_0029 43839 43390 CPT_U029 43839 43390
No robust homolog present in Genebank/EMBL as of 11/7/98
SNKNERNENIYCFNLFRYIRFFAALNIRMNDGLRFCYSYILLREMLLDSSLIRKGGQELL
KKFQIKLRTTSIKSSLISLRQQLGKREATQSDILYGTSRFQYLNSFEIEDPRIPPTMAAQ
LQEITWSRSVMELKIKFYVYLNSERNKTKP

43840 CPn_0030 44529 CPH_0030 43840 44529
gcp-0-Staloglycoprotein Endopeptidase
LKGYCWYSLFFYIKWRRMYFYKYVIIDTSGYYPFLACVDNQQVLEHWSLPVGPDLGIVLE
FLFKSKNLSFQGVAVALGPGNFSATRIGISFAQGLAMAKNVPLLGYSSLEGYLLSKDEKK
ALMLPLGKRGGVLTLSSEIPEEGLNEKRRGVGPGALLSYEEASDYCVAHGYYHVISPNPQ LFASSFSDKITVEEVAPSVEQIRRHVISQFMFVEYDKQLSPDYRSYSCIF

CPn_0031 44708 44884 rs21-S21 Ribosomal Protein CMPSVKVRVGEPVDRALRILKKKIDKEGILKAAKSHRFYDKPSVKKRAKSKAAAKYRSR

44881 46098 dnaJ-Heat Shock Protein J SLIGNVVFVGSVSGMDYYSILGISKTASAEEIKKAYRKLAVKYHPDKNPGDAAAEKRFKE

VSEAYEVLSDPQRDSYDRFGKDGPFACAGGFGGAGGMGNMEDALRTFMGAFGGEFGGGS FFDGLFGGLGEAFGMRSDPAGARQGASKKVHINLTFEEAAHGVEKELVVSGYKSCETCSG QGAVNPQGIKSCERCKGSGQVVQSRGFFSMASTCPECGGEGRIITDPCSSCRGQGRVKDK RSVHVHIPAGVDSGMRLKMEGYGDAGQNGAPSGDLYVFIDVESHPVFERRGDDLILELPI GFVDAALGMKKEIPTLLKTEGSCRLTTVPEGIQSGTILKVRNQGFPNVHGKGRGDLLVRIS VETPQNLSEEQKELLRTFASTEKAENFPKKRSFLDKIKGFFSDFTV

CPn_0033 46129 48171
pdhA&B/odbA&cdbB-(pyruvate) Oxolsovalerate Dehydrogenase Alpha
& Beta Fusion
ERSMGNUMDOULGGEREE & BECA FUSION
ERSMGVVONQVISSIRDVLKLVWELRFAEHKMLLLSRQSGSGTFQLSCAGHELAGVLAG
KSLIPGKDWSFPYYRDQGFPIGLGCDLSEIFASFLARTTPNHSSARMMPYHYSHKKLRIC
CQSSVVGTQFLQAAGRAWAVKHSSADEVVYVSGGDGATSQGEFHEMLNFVALHQLPLITV

CQSSVVGTOFLQAAGRAWAVKHSSADEVVYVSGGDGATSQGEFHEMLNFVALHQLPLITV
IQNNHWAISVPFEDQCGADLASLGRCHQGLAVYEVDGGNYTSLTETFSHAVDQARQHSVP
ALILIDVVRLSSHSNSDNQEKYRSALDLKLSMDKDPLILLEKEAINVFGLSPFEIEEIKA
EAQEEVRKSCEIAEALPFPSKGSTSHEVFSPYTETLIDYENSESAQNLRNSEPKVMRDAI
SEALVEEMTRDSGVIVFGEDVAGDKGGVFGVTRNLTEKFFDQNGFNSPLAEATIIGTAIG
MALDGIHRPVVEIDFADYIWFGINQLFSEASSIYYRSAGEWEVPLVIRAPSGGYIQGDPY
HSQSIEGFLAHCFGIKVAYPSNAADAKALLKAAIRDPNPVVFLEHKALYQRRIFSACPVF
SHDYVLPFGKAAIVHBXDLTIVSWGMPLVLSLEVAQELASRGISIEVIDLRTWPCDFA TVLKSLEKTGRLLVIHEASEFCGFGSELVATMSEQGYAYLDAPIRRLGGLHAPVPYSKVL ENEVLPHKESILQAAKSLAEF

CPn_0034 49496 48210
CT345 hypothetical protein
WNFLLPTTCRGILMAEISTPSLPDSSIVSQKTPPVPDPDSSPDHIPTIPTQAPFKPQRKK
ETPSSIVNAIAFALLAFLSCLGGFVAICLGCSLEITMPLFILTAVFIAFTLLYFIHYLEK
PKIPEPLFTPPPSPTLRAPTLTPEIPAPAPGIPLPPTLPKVDRTKLTCNPDIHYPSTYDP
KACFSLLKQLFSLDPETRPEDRKYSNKLASILLRSKEKSGFRFHCFKGHFSHDKILNKKS GAVVISSHSSMDFSTTLGRAFAVTTCLQRSCWEKIKNNIPTPEKHLPIGSCVSGPWDVEE GAQLYTSHLIVINPPTLETLIKEKMRRAITLKDFSMKEAFTNLVLAYLQCFDICIEHNLE SVQLEVFGLNNLSADQEEFTTWESCCHLALLESVRILLASKEEYALSNVSVNSISQVPLQ TACRALFLN

49569 CPD_0035 51146 49569
CT339 pypothetical protein
ARTTLEEDAGSSLKPLPKTPPCATALYITHRRERKSEHQMWNRCQVFSSFFFRYPISSWL
IRLRASCECFQORHPIFLCGLYWLAG ITSRGHPECSALILIFLGMFLPRNPKOWLPLASA
WIISLMLTPAPFLHDGPISGTFVIHHAGGGTTYYGEALCIQTFGGKRAHHLSCQILSESR
ELLKKVYELEGTLHHTSQIVFKSNACYKEIPRSFPYIMKEKCRESSGHFLNHRFPSSEVG
PFASSLLLGTPLPQNLRDLFRQKGLSHLFAISGWHFSLCATTLWMLCALLPLKIKKILSF
IVLTSLACIFFWSLSVWRSWISVTLLCFSWCFSGSCSGLNRLGAGFILCSIFFSPFSPTF
VLSFLATLGILLFFPKIFSFLYTPWTOFLSPFWLYPIRYLAMTLAISLAGLFIVLPIMQ
YFGSLPLEGLLYNLIVPFTILPIIVFLIATIILPCCSPITEALIQGFLSHPWLHNPNILK
TLSFAPVPPWMLTLASLILFFIGILRTNVSPYASISATCYRFIETL

CPIL_0036 50969 51796
CT338 hypothetical profein
AK:SUMDLERKMKKPDNDDFEDVRSFEPFDVLGTEQLPKEMSWEVVGAKTPPLPRGWYEL
MOLERCDRIDF LDDWC:VLGTEHKEOPSTCPFFSLLETTEVYTYPLEKEPYGLKMFYVF
RDGROGFGGEFGLDFGJHHRLPPLGDPHYEKFFSTHMSFOKWIDEGTFFMPGLKWQQK
LHCQUVMNNMOAEDNCYSIGTFFFYY/KEPFAYQGFFFDPETRPLLFSHMVLLNEGGE HROLDT LEPTHFORMANDELTMFENAFRAGERANE

51702 52115 CTEL_0037 - 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24127 | 24

53831 Fm 0038 52119

CFN_0038 hypothetical protein
MDTOSSTONEEWRIAGTSIVSGMALGKVFFLGTSPLHVRELTLPQEEVEHEIHRYYKALN
RSKSDIVALEQEVTGQQGLQEVSSILQAHLEIMKDPLLTEEVVNTIRKDRKNAEYVFSSV
MGKIEESLTAVRGMPJVVDRVQDIHDISNRVIGHLCQHKSSLGESDQNLIFSEELTPY
EVASANGAYIRGFVGLVGAATSHTAIVSRAKSIPYLANISEELWNIARFYNGKLVLIGGY
POBLIEBERGBERGATIDA PARTIN OF ART OF A RSIRWLLDYSVILEDOLDAIAKASLOGSIKVLIPGVSDVSEIIEVKKWETIRTRFPRGH KVSWOTMIEFPSAVWMIEEILDECDFLSICTNDLVQVTLGISRESALPKHLNVTLPPAVI RMIHHVLQAAKQNQVPVSICCEAACQLSLTPLFIGLGVQELSVAMPVINRLRNHIALLEL NSCLEITEALLQAKTCSEVEELLNRNNKITS

CPn_0039 54256 53963
CT339 hypothetical protein
ISMGSGYAKKKEAKIMEQQFLEMEASLLEKRYEGQAGNGLVSVVINGKCDLISVKVQPT
CLDPEDPEVIEDLFRAAFKLAKEQMDQEMSLMRSTMPF

55673 54318

CPn_0040 55673 54318
dnaX-DNA Pol III Gamma and Tau
AFYTHSLGYTMTLOPYQASSRKYRPQIFREILGQSSVVAVLKNALVFNRAAHAYLFSGIR
GTGKTTLARILAKALNCVHLSEDGEPCNQCFSCKEIASGSSLDVLEIDGASHRGIEDIRQ
INETVLFTPVKAFKIYIIDEVHMLTKAFNALLKTLEEPPQHVKFFFATTEIHKIFGTI
LSRCQKMHLQRIPEKTILEKLSHAQDDHIEASQEALEPJARAAQCSLRDAESLYDVIS
LFPKSLSPDTVAQALGFASQDSLRTLDNAILQRDYATALGIVTDFLNSGVAPVTFLHDLT
LFYKNILLINSTTSKFSSQYKTEQLLEIIDFLGESAKHLONTIFEGTFLETVIHIIRIY ORPVLSELISSIKSROFEGLRNIKEPTLTOOVSAPOPOPTYKEOSFLEKKNOPAAEGKII SVEVKSSASIKSAAVDTLLQFAVVEFSGILRO

55888 57342

CPn_0041 55888 57342

No robust homolog present in Genebank/EMBL as of 11/7/98

CKYLYHSYPPPCHSVGSISSRYKLRVLAITFLVLGVLLLISGALFLTLGIPGLTAGVSF
GLGIGLSALGGVLVVSGLLCLLVKREVSKVCPEEIPAVQPEETPBGVPVTPFEKPALDEA

GKEQKTQKILDQLPQELDQLDRYIQEAFACLGPLKDLKYEDQGFLQDVKEEFQVFDFVQK

DMIAEFVELQQILCQEGRLLEFVINQTRYIGRDLFKREDSLYKLWEWLGYLPSGDVRGER

LKKSAREVVDRFMRTTCNIRKIAMTFDRHVYSVAKTAFEKAFGALETCVYESMRESYREA

FCEYEKAKLLGDEKSSAHAEORFODIKNWEDVKDAFFWYMEDGKIEIDDAIGNSCWSE

RYEEHRITRARWYKVAEHQLFNATMRVKDSLREHNEARVAFEKERSKENQRQVQKKKEKR

LRÖBKELHDQELPRAQERLRELQALYPEIAVSVVEARREVASDLEKAHESIDKHYQSCVR
EÖEEY

58432 60372

CPT_0043 58432 60372

NO_CODUST homolog present in Genebank/EMBL as of 11/7/98

HHRTIMOVPLSPQLPPPPDHSVGASFCLSKFRVLAITFLVLGVLLLISGALFLTLGISG
VSLGVGLGLSALGSVLVISGFLLLLERREVSGVGLEGIPTGIPVGPSAEPSSEEIQKKOK
AKQILDQLPQELDQLDTDIQHVLSCLGKLKDLKCKDRGLLKDAKEKLQVPDFVWKDMMME
FVELQQVMDQESRYLEGLIHEVQSIAHKLFVDDVINIRSHLGESGYLPSEDDVGGELLKRF
VFLQQVMDQESRYLEGLIHEVQSIAHKLFVDDVINIRSHLGESGYLPSEDDVGGELLKRF
KREVVARFMKVTRDIRKIAMAFNKNAYGAAKNAFDKAFGSLETCLYKSLTKSYRDTPCDY
KRAKILPDENNSARAEQFFREVKDHWEDLKETVFWVKEDGRIDIEVLTAVGGWPDRYPEH
LILEKRKDKVMSHQLWEATMRVKEAEVTYSVARVAFEKDGSQQNQKKRQEKTKERLRCLK
DLRDGCKRAQGRLEKLTALYPEVSVSVVETERERKFNLEKAYGNLEEPYGSVVQDEDY
WTEQKNREAEFRAKGTKVRSMEEVAEHLQILENLLEDCYKRLSKAETFALGVEREATEEI
EYT ILSDAANRLKVLCEDIEDTLPRVEEIEMLRMAERPLHPIKQAFTKAFVQYNRCKER
LAKVEPYYKESPAYVNSEERLQSLDQASQCIORVPKGFKFRNGSMYI

CPH-0044 60278 60778
NOLEDBUST homolog present in Genebank/EMBL as of 11/7/98
IAKSDCRVWIRLHSAYKESQKVSSLETEACTYREYLREQVQOFETQGVSLIKEELLFLSS
TLKSKLSVDPLIANIPCMKFYYQYYDDIDKARAQSRWLEKSERYRNAKRFQEIVKKGLF KEAKPLKKEEYRLLQEERSNKEKRLIYNKMAVARQRVQEFESMEIPE

CPn_0045 60961 62790

CPn_0045 60961 62790
CT345 hypothetical protein
CKYTYHPPOLPPDHSVGATSMOPKLRILTITFLVLGVLLLISGALFLTLGVPGLAAGLSF
GLGIGLSALGGVLVVSGLLFFLIRRGVSKVRPEEIPVTPSHEAQKILCQLPOELDOLDTS
IQEVVSCLGKLKNDLKYEDQGLITEVQEKLRVFDFVRKDMVTEFLELQQVVAQEGQFLDYL
INQVQS ISHKLFVPDVDNIGAHLAELCGYLPSGDVRVERLKRSARQVVDRFMRVTCDTRKV
AMAFDENACGVAKNAFDKAFGALEECVYKSLTESYPEAFYEYEKAKILRNEDVEMLQDKN
KSARAEQRFREVKDRWEDLKETVFWVKENGCIDLEVLTAVCGWPDRGPEHLIPEKRNKV
MSHKLWEAATMRMKGAEGTYSVARVAFEKDGSRKNQKKFQEKTKEWLRCLKDLHDQECHRA
RERLAELEALYPEVSVSVVETERETKFKLETAYGNLEERYGSVVRDQEDYWKEEDKKEAE
FREKGTKVRSPEEVVEYLQILENLSEDCSKQLTIAE*/VLGVELEATAEFEYTILSDAAN
RLKVLCEDIEDILPRVEEIEIMLRIAELPFLPIKQAFTKAFLQYNSCKDKLAKVEPYCQE
SVDYKSGFRV SVDYKSGFRV

CPn_0046

CPn_0046 62775 63263
No robust homolog present in Genebank/EMBL as of 11/7/98
ERFQSLNQDLQNVYQECQKATGLESEVSAYRDHLREQITEFETQGLDVIKEELLFVSSTL
KSKLSYDPLIADIPCMKFYEEYYDGIDKARVQSRWLEXSERYRKAKKGFQEMLKEGLFKE
DQALKKAEYRLLREKRMNKEKLLICNKIEAAQQRVQEFGPSDS

CFm_0047 63437 63652
No cobust homolog present in Genebank/EMBL as of 11/7/98
EHFFLEKYTIVEFFREVLEGILTMYHFOKIPMTLTTYSFYLNKSLRKDYELWFVYGGCPES KVKLQTOSHEWL

CET_0048 63687 65801
*YGTE BG CORRECTED BYPOCHOET AT IM DEFFECT
MKELPHEZYNRALHKUSHOWWRYFLYTEVOCHEUVALITERAWLKVLYVPEYKAGEISRIS
LYDARMOETUJWCJAHKFYKRTAHICBAFCKVYHLDIJFJGLGKECNADENTOWFKKAAD
FOLDTNEVDGTOKCUKDLCIYPPLEGKEKKTLEITHICKNKGNVIAQCFCHLKIFLIQEN

CPOPCEDA IMDILKIANFEVAVOKEMSGCVKGELLGKRCIEKITKGTPILEKYGRIDDRD AKILKQLRAQLLSVHTLFSCRSLWGA IFVVLLILLWGYGALKALCPEMLKSPQRFMLYIA ILTLSLLWCPGTEIFCAYWVYLGYPPILPFTAVLLGYFLGLPIAGFSCTFLALLYTLGS DLWNNSWFLSINLLCSWRILVSLHRVSRLSSVFWACMKLGGVANGSLLMFRIFTNTISRE ALYADGIESFVYSLITAISVVALIFVFEASFGASTNFSLLTYLSPENALLKRIFKENGT YQHSVLVGSLAEAAAQAIGADSLVCLVAAHYHDIGKLINFGFFSENQKILQQSGHSLSPL ECAKMIMRHIPBGVNLARQAGLPESFIQVIEEHHGTSVIRSAYYSHMVENPSTGSFDEEL FRYSGNKPSKETTIIMIADSFEAASRSLKNASLPDLQRLIDQIIQGKLQDGGFSCSPIT

00817

CIM_U049 66362 63817
NO robust homolog present in Genebank/EMBL as of 11/7/98
LKEKRRNIVYLLVIYQEIFWLTMLHQPYYDKILTGNTIYIPGHTHKDSNKLFQKKSRAIW
VDEKPFSLDCFSNVFLIFVSLVPIAGLVRAYQIKKSLDRTTVQIGYSPSLSCEQKECVEA
FVXXYGLICISILGGLGILVPILILVVLSLLLLGILMLFSLSTYESIKNYISKHICWKSN

CPn_0050 66849 66499
No robust homolog present in Genebank/EMBL as of 11/7/98
vswFPILGIFLAMRYAKHOTNWNDENVKANLGYLPSTNCKNALCRMSSRLTSSIKTAGIL
GGCGILLPIFLLLAILLISVLFQLIMLPFRLCCFALRQSVSSDTVTNLLLLNNTLA

CPn_0051 66797 67111
No robust homolog present in Genebank/EMBL as of 11/7/98
CFAYLIARNIPRMGNHETYIHPGVLPSSHAQDVSRSTVYPSRSFIMRRMLMGWNFNRVPS
KSSEQLMIGHRIPLIFFGKHHPIISILMVNRFSWLSIFYNGERGF

67304 68008

hemC-Porphobilinogen Deaminase
kMLSVCYSDPCLSDFCQGKRPLRIASRNSNLAKAQVHECISLLRSWYPKLWFQLSTTETT
GDREKKIP.HLVENSYFTDGVALVHKGVCDLAIHSAKDLPETPSLPVVAITRCLHPAD
LLVYADHYVHEPLPLSPRLGSSSLRRSAVLKQLFPQGQILDIRGTIEERLDQLHRGHYDA IVLAKAASLRLHLHHAYSIELPPPYHALQGSLAITAKDHAGKWKQLFTPIHCHSS

69350 67986

sms-Sms Protein

sms-Sms Protein
IRMARKTKTOWTCNQCGATAPKWLGQCPGCHNWNSLVEEYVPQARSGTSSRSSTSAIALS
SIELENESRIFIDHAGWDRILGGGVVRGSLTLLGGDPGIGKSTLLLQTAERLASQKYKVL
YVCGEESVTQTSLRAKRLNISSPLIYLFPETNLDNIKQQIATLEPDILIIDSIQIIFNPT
LNSAPGSVAQVREVTYELMQIAKSAQIITFIIGHVTKSGEIAGPRVLEHLVDTVLYFEGN
SHANYRMIRSVKNRFGPTNELLILSMHADGLKEVSNPSGLFLQEKTGPTTGSMIIPIIEG
SGALLIELQALVSSSPFANEVRKTAGFDPNRFSLLLAVLEKRAQVKLFTMDVFLSTTGGL
KIIEPAADLGALLAVASSLYNRLLPNNSIVIGEVGLGGEIRHVAHLERRIKEGKLMGFEG
AILPEGQISSLPKEIRENFRLQGVKTIKDAIRLLL

70089 69313

rnc-Ribonuclease III

TIC-ALDORIUC BEASE III
TLSFFPPIKIPNSKFKDGALLSMHPPIDITAIEAKLNFTFTOPKLLEIALTHPSYKNESA
VQIEDSERLEFLGDAVLGLIVTEHLFLLFPSMDEGTLSTARASLVNAKACCRYTTMLGIG
DYLLIGKGEKIQSERGRLSAYANLFESILGAVYLDGGLSPARKLTVPLLPPREEILPLMS
GNPKNLLQOFTOKOFFRUPVYQSTAVTDAQGNVSYQIQVLVNQEVWGEGNASSKKEAEKI
AAQQALDTYGNKNQNTMDV

CPn_0055 70096 70590

CF120039
CT296 hypothetical protein
CFWICYLIRIRMRSALHLOHLRHFHNHGSILFENLLTIKDCFLLETKLONFIAKASKTID
TYRMRENIFRSMPEITYUVRKRRLDFFAAELUHRPKLSLVRDLWVFPGEEILEGEEDCML
FLLLSGDRAGSGIFFTGPYPSDLYELEKGTTGLLLAFSSVGIPVI

70917 72746

CPn_0056 70917 72746
mrsa-Phosphomannomutase
EFLKLSLHRISLMKEVEORIRSLYDAVTAENICRWLSNDCTQQDAKTILGWLDTDPAQLE
DLFGATLTFGTGGLRSLMGIGTNRINLFTIRRTTQGLVQVLRAHLPHPGDPMRVVVGCDT
RHNSIEFAQETAKVLAGNGCEVFLFQYPEPLALVSFTVRYERAIGGVMITASHNPPNYNG
YKVYMASGGVLPPLDQEIVAACSAVNEILSVPSIDHPNIHLIGKEYEALYRDTLKOLQL
YPEANRISGRSLSISYSPLHGTGISLVPHVLKDWGFLSVHLVEKQAIGDGDFPTVQLPNP
EDPEALTLGIEGMLANDDDLFIATDPDADRVGVVCLEDGGPYRFNGNGMASLLADHILGA
WSKTRHLGEHDKLVKSLVTTFMLSAIAKHYHVDLINVGTGFKYIGEKLESWRNSTNKFVF
GAEESYGCLYGTHVEDKDAIIASALIAEAALQQKLQGKTLCDALLSLYETYGYFANKTES
VVFSAKTDEQEIRKKLSHLEEISSANFFSGKYQVEKFENYKQGIGFNLLSKDSYALTLPK
TSMLCYYFSGGGRVIIRPSGTEPKIKFYFEMSTHYPERVTDKEIQKQREAESFQHLDDFI
FDFKEKFSNL

73554

sodM-Superoxide Dismutase (Mn) SOUM-SUPERINGE DISMULASE (MI)
LIKRYVVMSFVPYSLEPLPYDYDALEPVISSEIMILHHQKHHQIYINNLNAALKRLDAAE
TQONLNELIALEPALRFNGGGHINHSLEWETLAPIDQGGGOPKHELLSLIERRWGTMDN
FLKKLIEVAAGVOGSWAMLGFCPAKQELVLQATANQDPLEPLTGKLPLLGVDVWEHAYY
LQYKNVPMDYLKAFPQIINWGHIENRFSEIISSK

CPn_0058 73627 74562
accD-AcCOA Carboxylase/Transferase Beta
IRWLVRLFSYDKPKIKVQKIKADGFSGMLKCNHCHEMIHANELGQNYNCCPKCSYHYRIT
AIERVKLLADKDSWRPLYTDLKSQDPLEFIDTDTYANRLEKARKNTTESEGVIVGICTIG
LHPVALAVMDFNFMAGSMGAVVGEKLTPLIEEAIETRLPVIIVSASGGARMQESVFSLMQ
MVKNSAALAKLHEAGLPYISVLTNPTSGGVTASFAALGDIIIAEPKALICFACPRVVAQV
IGEDLPEGAQKSEFLLEHGMIDKIVERKELKTTLOTLLDYPLAQEYTGGKSKAPRDLSKR

CPn_0059 74562 75050
dut=dUTP Nucleoridohydrolase
KKHHTASCNDNIICNAILMTVFCELDSSGELPEYTTFGAAGADLRANIEEPIALLFXXXRA
LIPTSIKAEIPESYELQVRRRSGELALFHGITVLNSFCTIDSDYRGEIRVILLINFGDSTFI
IEPKMFIAQVVLGPVVQATFVVKQEGLAETARGSSGFGHTGAS

CPN_0050 75004 75528
ptsN-PTS_TIA_Protein
PKLPEEYEVIV LEQARMESYCQNOQDESLESSLSPRLVMFLGKHSRDEILQDLTDLVDA
AGLLEDYQAFFDALVRRENIMSTSIGM:7VAIPHGKLESSSNFFIAIGHFQAILWHAIDG
ALVRLVFLIGGPENAQAEYLKLSTSTTLGTRESPROQULQVNTIEEVMNVFVGM

CPn_0061 75501 76209
ptsN-PTS [IA Protein + HTH DNA-Binding Domain
RSHECICGOVKMDLKLDEVASLLDVSGEHTVLQWLKEGAIPSYSMNNEYRFSREEIENWLL
HNQALMIQERGEDKEALKDLSLKYSLYKAIHRGGVLCDVVVHSKEEALQYASKYIAQKFQ
LDESVLFEMLSHPENLMSTGIGECIALPHAKDFLINAYYDIVVPMFLAEPIEYGALDGKP
VGILFFLFACQDKSHLNLVNKIVHLGMSLNARSFFKNYPNKDQLLAYVKEWESQTH

CPn_0062 76251 77690

78267 78109

No robust homolog present in Genebank/EMBL as of 11/7/98 PMYANCKHNCLCLYDFSRHRSPPGLPLTFTPPYSFTLGIFLGRCLSTSNIVLL

CPn_0064 78340 78576
No robust homolog present in Genebank/EMBL as of 11/7/98
LVMTKIQCSAQYYRSRPAERAQTPPQPFLARDRADFWERHPRFSACCRVLLLVAWVVLAL
LFLFVMLLPLAAGSYLLAF

78882 80651

CT1288 hypothetical protein
YDYYKYNNFFKKNYMTDFPTHFKGPKLNPIKVNPNFFERNPKVARVLQITAVVLGIIALL
SGIVLIIGTPLGAPISMILGGCLLASGGALFVGGTIATILQARNSYKKAVNQKKLSEPLM
ERPELKALDYSLDLKEVWDLHHSVVKHLKKLDLNLSKTQREVLNQIKIDDEGPSLGECAA ERPELKALDYSLDLKEVMDLHHSVVKHLKKLDLNLSKTQREVLNQIKIDDEGPSLGECAA
MISENYDACLKMLAYREELLKEQTQYQETRFNQNLTHRNKVLLSILSRITDNISKAGGVF
SLKFSTLSSRMSRIHTTTTVILALSAVVSVMVVAALIEGGILALPILLAVAISAGVIVTG
LSYLVRQILSNTKRNRQDFYKDFVKNVDIELLNQTVTLQRFLFEMLKGVLKEEREVSLEG
QDWYTQYITNAPIEKRLIEEIRVTYKEIDAQTKKMKTDLEFLENEVRSGRLSVASPSEDP
SETPIFTGCKEFAKLRRQTSQNISTIYGPDNENIDPEFSLPMMPKKEEEIDHSLEPVTKL
EPGSREELLLVEGVNPTLRELNMRIALLQQLSSVRKWRHPRGEHYGNVIYSDTELDRIQ
MLEGAFYNHLREAQEEITQSLGDLVDIQNRILGIIVEGDSDSRTEEEPQE

80916 82655

GPN_0066 80916 82655

NO-TODUST hOMOOLOG PIESENT IN GENEDANK/EMBL AS OF 11/7/98

GYMANPTOSRPPSPEISIEELELQELAGSSNTETISNTPPPSCAATAEEVSLFIEGGRR
NEDEEGPLGSCEVYDVVCITNQGDPEVRDHEVRVMYINGSGRTQHEGILDAMNICDLRG
EEVRFIHNSGYGLGSCFLGIRNRIPPRDNVISQAIQARWNEFFIFAENANRDYIVLFSGN
GELYLQVALDNSIYSHHILCVGIGSSYYIQGNYRVHNYRVTGDWTTLLDRRGATAVNTTT
LEPADSAEGLFLPSVRCPSYGWALRCGEQCLIMDNNQQVGFRPQDSSSEIALVWLINQDH
STÄTTLIEWIDRGDSQAVLEINPOPSHCRDIALTALYATTRISSLLQECLMISVITYAPEV
FATYAIVTGYSIMTLRYFILLLTNRPGCRHFRVLRLAALGLQSLGFLTVLLDHINVTRR
WRRPPLISVIFCTASFATGSFIYVDLTRMFFTSLRSRLQLFVQRRLTGRGLPLRRVFVN
HÜDSLRFSQNALITFHGGLFMPLIIGFFNQLVIQVRRVVIRPNTTAVYDLNQTSQEAWDS
GEVLAIGQTINFLLCMILLVINTFFFVRSVRRNLHRRPHR

CPHE 0067 82920 84053
NÖ TODUST HOMOOLOG PIESENT IN GENEDANK/EMBL as of 11/7/98
KGSGYSYRGPMAVEGRVNSSQALNODCQEVLANKQSKGLLRCRILSIVVAVITFIAGVV
LIALTLASILTSVPYLALGVFLLIVTLGCIFFALCSEKIKKVPPPPISHKEE ILAWFEER LIADILASI LISVPYLALGYPLLIVILGCI IFALCSEKI KKVPFPFISHKEE I LAWFEEK
KRÉDMEKEKEDPEHFERTATDI PMRSALDOFNISCHHI HESPALITETYRSHQDVLLFKDW
CPVILPDVTSEEEVLI RSVVGSYLLMEACVPKVSML I DELHNKLKSPSERECLFI DKKTL
QRKASFLFTQKDLATFFLAYTRVNIGHLAPFRAGAKWILLIHYVRLRRQHNQNDFFTEGHS
CYYARLAFNQTQRLYHQLFNVEKLRSIYANMDKDPLCHPWAFI PIYDLLKTEDHGDGFLE
QQEDREYPSRAAQDQFWG

CPR_0068 84909 84331
CTF60 hypothetical protein
SFMIKKFFIYSLIFSCSFSAPLKGICNEDVSSQSRIEEDPEVLITQLNELIETPIEEGKE
IRRELQAISDGQKSSEEIEESCGTSDSEGLSEKTDKESSNEYVLDFFDSMVQRLEGISKM
CQSGQVAQIIDCFNREFDIRNRELELKNRELELREKDLEFKKSILDWNKEKVSRELAFQR

85191 87086

CPn_0069 85191 87086

No robust homolog present in Genebank/EMBL as of 11/7/98

No robust homolog present in Genebank/EMBL as of 11/7/98

No robust homolog present in Genebank/EMBL as of 11/7/98

No robust homolog present in Genebank/EMBL as of 11/7/98

No robust homolog present in Genebank/EMBL as of 11/7/98

No robust homolog present in Genebank/EMBL as of 11/7/98

PYFSASNFVVIERGAPSEQULISLPEYSROPPPGYFDETSITSRTSEEMFGTLVSTAGE

ENCHAPPEYSRAMTORIVATARATY FWDRFFSRHWNVGRFFLVFYQQNGGAYVQAALDSSMHTQDI

YVLGLSPTVYIRONYHVQHYRVRGFWPSCLDSLAACAENTSVLPYGESSGIFYPSLFSH

TPDNAIRYGERCLLVCSEOMGMLPETQQQTSPLTSLEGGHEVALVLNPQQNPEALSIASR

LMHEERGGRLESNYMPGRSSNPFMTSMYVLVRLNTLAQIYLMSPYYSTQSNDIVCLIFIS

SAAVETVSYIFLTVTDSTCGRRYLRVPRLVCTGLRNLALPTTLLELLILSYRRSVEGVPF

NVRFILGYMCTTRVVFFAWNLILHHPFRCLRHGIGLFVHRSIIGHTLGARITDLTLASMR

YAIVPPSIVSSCLLTALAHANTNILALDPYRLIESGDLRRPAFNDDEMQQADNPWDAYSI

GLVINTCIYMLILFANLIFMVYSVRRYHRSRR

CPn_0070 87399 87208

No robust homolog present in Genebank/EMBL as of 11/7/98
YKVGLFHLKNONFFSNQSRTYEQRFPKVSPHFESILPLQSVGFSSQGTLLISFRDTELKR

CPn_0071 88066 87599
'TT:35 hypothetical protein
KKOLRG LEFTCPLOHARCLKKOHKI IEELFPEPFOKDHLYLKLMENSSSRDAFDKKRML
KENLJVGCQGDLYLYEVYQDG LEFFTYTKALMDDG IASLFTEVYSGETPSTILTCKPIF FORUTPYLEFORLNOGESLYMRMKOIAVQYLKPPQT

89151 88057

CTH_UOT:

##UST
##

eatyaidpkahkkpienieoainchqiikhqminokolnalieinrnnctdpatanllas Lklnlnopmpycfompecgvtgsyldlnnnopddiiaradocimtlsctlooikkepdri LESNH

83353 CPn 0073 89574

infA-Initiation Factor IF-1

SMAKKEDTLVLEGKVEELLPGMHFRVILENGMPVTAHLCGKMRMSNIRLLVGDRVTVEMS AYDLTKARVVYRHR

84762 111/122

tutA-Elongation Factor Tu

EUIA-ELONGALION FACCOT TU
EDFEMSKEFFQRNKPHINIGTIGHVDHOKTTLTAAITRALSODGLASFRDYSSIDNTPEE
KARGITINASHVEYETPNRHYAHVDCPGHADYVKNMITGAAQMDGAILVVSATDGAMPQT
KEHILLARQVGVPYIVVFLNKVDMISQEDAELIDLVEMELSELLEEKGYKGCPIIRGSAL
KALEGDANYIEKVRELMQAVDDNIPTPEREIDKPFLMPIEDVFSISGRGTVVTGRIERGI
VKVSDKVQLVGLGETKETIVTGVEMFRKELPEGRAGENVGLLLRGIGKNDVERGMVVCQP
NSVKPHTKEKSAVYVLQKEEGGRIKPFFSGYRPQFFFRTTDVTGVVTLPEGTEMVMPGDN VELDVELIGTVALEEGMRFAIREGGRTIGAGTISKINA

91087 91350

SREE-preprotein translocase
SRSWFMKQQHNRKALSRKIGTVKKQAKFAGSFLDEIKKIEWVSKHDLKKYIKVVLISIFG
FGFAIYFVDLVLRKSITCLDGITTFLFG

91903

nusG-Transcriptional Antitermination
OPFCSVNCMYKWYVVQVFTAQEKKVKKALEDFKESSGMTDFIQEIILPIENVMEVKKGEH
KVVEKYIWPGYLLVKMHLTDESWLYVKSTAGIVEFLGGGVPVALSEDEVRSILTDIEEKK SGVVQKHQFEVGSRVKINDGVFVNFIGMVSEVFHDKGRLSVMVSIFGRETRVDDLEFWQV EEVAPGQESE

91956 92435

Till-Lil Ribosomal Protein
FFVSYPLFVEVSQCKVRFSMSVKKVIKIIKLQIPGGKANPAPPIGPALGAAGVNIMGFCK
EFNAATQBKPGDLLPVVITVYADKTFTFITKQPPVSSLIKKTLNLESGSKIPNRNKVGKL
TQAQVEAIAEQKMKDMDIVLLESAKRMVEGTARSMSIDVE

92453

T11-LI Ribosomal Protein
SCRIMTKHGKRIRGILKNYDFSKSYSLREAIDILKQCPPVRFDQTVDVSIKLGIDPKKSD
QQIRGAVFLPNGTGKTLRILVFASGNKVKEAVEAGADFMGSDDLVEKIKSGWLEFDVAVA
TPDMMREVGKLGKVLGPRNIMPTEKTGTVTTDVAKAISELRKGKIEFKADRAGYCNVGVG
KLSFESSQIKENIEALSSALIKAKPPAAKGQYLVSFTISSTMGPGISIDTRELMAS

93688 93170

Tilo-Lio Ribosomal Protein
RGKMKQEKTILLQEVEDKISAAQOFILLRYLRFTAAYSREFRNSLSGVSAEFEVLKKRIF
FKATEAAGLEVDCSDTDGHLGVVPSCGDPVSAAKQVLDFNKQHKDSLVFLAGRMDNASLS
GAEVEAVAKLPSLKELRQQVVGLFAAPMSQVVGIMNSVLSGVISCVDQKAGKN

CPn_0080 93720 94121
rl7-L7/L12 Ribosomal Protein
VRVTKVTTESLETLVEKLSNLTVLELSQLKKLLEEKWDVTASAPVVAVAAGGGGEAPVAA
EPTEFAVTLEDVPADKKIGVLKVVREVTGLALKEAKEMTEGLPKTVKEKTSKSDAEDTVK KLODAGAKASFKGL

CPn_0081 94219 98016

rpoB-RNA Polymerase Beta
FREILSHONSRRTRMLKCPERVSVKKKEDIPDLPNLIEIQIKSYKOFLQIGKLAEERENI
GLEEVFREIFFIKSYNEATVLEYLSYNLGVPKYSPEECIRRGITYSVTLKVRFRLTDETG
IKEEEVYMCTIFLMTDKGTFIINGAERVVVSQVHRSPGINFEQEKHSKGNILFSFRIIFY
RGSMLEATFDINDLIVHIDRKKRRKILLAIFIFIRALGYSSDADIIEFFFTIGESSLRSE
KDFALLVCRILADNIIDEASSLVYGKAGEKLSTAMLKRMLDAGIASVKIAVDADENHPII KOFALLVGRILADNI IDEASSLVYGKAGEKLSTAMLKRMLDAGIASVKIAVDADENHPII
KMLAKDPTDSYEAALKOFYRRLREGEPATLANARSTIMRLFFDPKRYNLGRUGRYKLINRK
LGFSIDDEALSQVTLRREDVIGALKYLIRLKMGDEKACVDDIDHLANRRVRSVGELIQNQ
CRSGLARMEKIVRERMNLFDFSSDTLTFGKVVSAKGLASVLKDFFGRSQLSGFMQTNPV
AELTHKRRLSALGPGGLNRERAGFEVRDVHASHYGRICPIETPEDPNIGLITSLSSFAKI
NEFGFIETPFRIVRDGIVTDEIETWTADVEBECVIAQASASLDEYNMFTEPVCWWYYAGE
AFEADTSTVTHMDVSPKQLVSIVIGLIPFLEHDDANRALMGSNMQRQAVPLLKTEAFVVG
TGLECRAAKDSGAIVVAEEDGVVDFVDGYKVVVAAKHNPTIKRTYHLKKFLRSNSGTCIN
GQPLCAVGDVITKGDVIADOPATDORGELALGKNVLVAFMEWYGYNFEDAIISEKLIRED
AYTSIYIEEFELTARDTKLGKEEITRDIPNVSDEVLANLGEDGIIRIGAEVKPGDILVGK
ITPKSETELAPEERLLRAIFGEKAADVKDASLTVPPGTEGVVMDVKVFSRKDRLSKSDDE
LVEEAVHLKDLOKGYKNQVATLKTEYPEKLGALLLNEKAPAAIHRRTRAEIVVHEGLLFD
QETIERIEQEDLVDLLMPNCEMYEVLKGLLSDYETALQRLEINYKTEVEHIREGDADLDH
GVIRQVKVYVASKRKLQVGDKMAGRHGNKGVVSKIVPEADMPYLSNGETVQMILNPLGY
SRMNLGQVLETHLGYAAKTAGIYVKTEVFEGFPEGRIWDMMIEQGLEEDGKSFLVDGKTG
ERFDNKVVIGYIYMLKLSHLIADKIHARSIGPYSLVTQQPLGGKAQMGQRFGEMEWWAL
EAYGVAHMLQEILTVKSDDVSGRTRIYESIVKGENLLRSGTPESFNVLIKEMQGLGLDVR
PMVVDA

97992 102221

CPR_U082 9/992 102221

rpoC-RNA Polymerase Beta:
CSSYGRRLKNDVLEKIMFGENGRDIGVLSKEGLFDKLEIGIASDITIRDKWSCGEIKKP
ETINYRTEKPEKGLFCEKIFGPTKDWECCCGKYKKIKHKGIVCDRCGVEVTLSKVRRER
MAHIELAVPIVHIFFKTTPSRIGNVLGMTAGDLERVIYYEEYVVIDEGKTDLTKKQLLN MAHIELAVE IVHIWFEKTTESR IGNVLGMTAGDLERVITYEETVVI DEGKTOLTKKOLLN DAQYREVVEKWÖNDAFVAKMOGEA IYDILKSEDLQSILKDLKERLRKTKSQOARMKLAKR KI IEOFVSSNIHEEMVLKNIE PVYPEDLER PLVPLDOGGFATSDLINDLYRRVINNINRLK ALLRIKTEOFVSSNIHEEMVLKNIE PVYPEDLER PLVPLDOGGFATSDLINDLYRRVINNINRLK ALLRIKTEOFVSSNIHEEMVLKNIE PVYPEDLER PLVPLDOGGFENDLEKKKOGSVVTIRSAKKM ORGAREVWDVLEEI IKGHEVLINPAPTLHE DEI TOAFEPVLIEKKA IRTHELVGAAFNAD EDGDOMAVHVPLSVEAQLEAKVLMMAPDNIFLEDGGREVA I PSKOMTTIGLYVLMADETYF EFFHORKTIF IFORE EVLRALINNIYS I DDVPCIDER DETTROIH HEK IKVR I DOQI I ETT PGRVLENR IVPKELGFONYSMPCPF I DILLIOGYKKVGLEATVIRELDDLKUR I DOQI I ETT PGRVLENR I VPKELGFONYSMPCPF I DILLIOGYKKVGLEATVIRELDDLKDGT I OATKA ALJWILGKOTRINELELMI DILMOTADFORKOLLOGGGREKOLMARENGA I EFF I ETGNERE GLTVLEYG I SCHGARKOLADTALKTADOGYLTPPLVDVAQDV I TEKKNGTLINHI EI SAI OOGIGEELLPLKDR I YGRTVAEDVYOF JOKGHLLAOGGDVLNSVQAEA I DOAGIET IK IRS TELTVICJERICKOCKKAVILLANDE I JEMBER OGGIGEELLPKDR I TRAGTI TAGTI TAGTI TORGERTOLTRITETI I TAGTI TAGTI TAGTI TORGERTOLTRITETI I TAGTI TAGTI

NKNTGLVEL IVKQHRGELHPQIA IYDDADLSELVGTYA I PSGA I ISVEEGORVDPGMLLA RLPRGAIKTKDITGGLPRVAELVEARKPEDAADIAKIDGVVDFKGIQKNKRILVVCDEMT GMEEEHLIPLTKHLIVQRGDSVIKGQQLTDGLVVPHEILEIGGVRELQKYLVNEVQEVYR LOGYDINDRHIEITVROMLOKVRITDPGDTTLLFGEDVNKKEFYEENRRTEEDGGKPAQA VPVLLGITKAGLGTESFISAASFODTTRVLTDAACCSKTDYLLGFKENVIMGHMIPGGTG FETHKR I KQYLEKEQEDLVFDFVSETECVC

102296 103312 TELEGRAPH NOTEGRATATIVE CONTRACT

QELLNEAVVWGIRQNGDDLQTLSFILDKIQVNFALEIIKNIPGRISLEIDARLSFNVEAM VQRAVFLSQLFEAMGGDKKRLLVKIPGTWEGIRAVEFLEAKGIACNVTLIFNLVQAIAAA KAKATLISPFVGRIYDWWIAAYGDEGYSIDADPGVASVSNIYAYYKKFGIPTQIMAASFR TKEQVLALAGCDLLTISPKLLDELKKSQHPVKKELDPAEAKKLDVQPIELTESFFRFLMN EDAMATEKLAEGIRIFAGDTQILETAITEFIKQIAAEGA

CPn_0084 103356 103751

Dredicted ferredoxin
SEMKNKMDYKSQLVFSCPCCCKGNVCFSVFNLDVILTCNVCSSTYTFDSVIRNEIRQFVA
LCKRIHDANSILGNATVSVSVEDNQMDIPFQLLFSRFPVVLNLSLDGKKIAIRFLFDALN TSILHOESDLIS

CPn 0085 104512 103766

CT311 hypothetical protein
FSMKFFILFILIVAQFPAFSAQPRTQVSASHSKQAKARRTSRIRSSAATNASVSRYKTRA
AARKKIGKFEKKPSLSPVQWVRYSGKNYSIQTPSLWQCIDDKTQLPEKLDVLLIGKGKGN
LTPTINIAQEITSKSSKEYIEEILAYHKANEMTLESGIFTQIQSPSGEFTIIKTEKNSSW GRVFCLQATTVIDHTAYIFTSTATLDDYAELSFTFLKVVSSFQIRGGKEATSGDAILEKA LEALQNENK

104898 105527

depe-atp Synthase Subunit E NIMANINADGKIKQICDALRIDTIKPAEDEAAALLHNAKEQAKRIIQEAQEEARKILETA EERAHQKIKQGEVALSQAGKRALEALKQAVENKIFRESLVEWLEHVTTDPEVSTKLIQAL VQALEAQGVSGNLTAYIGKHVSPRAVNELLGKAVTTKIRKKSVVVGSFVGGVQLKVEEKN WVLDLSSSALLEIFTRYLOKDFREMIFOGS

105540 106376

CFI_U08/ CT309 hypothetical protein SHEKIFSIFKVVVMTQYYFLSSFLPTQLPESVPLFSISDLDDLLYLNLSENDLCNYGLLK RFFDFENFAFFWAGKFIPFSFGEVTQENVERMLSSQQWSDDNDFEDFFKDFLMNHKSSQD KINHFSDLFREFLSYHQTNSSKFLQDYFFFQQQLRVVLAGFRARVLNMDVSYVLRDEDSS DEWLEVLMQKDSPNYELPEEFSDLQGVLDDYGLLPHTLNRALALYQFHKLEGFCSDSYF DEWLEVLMQKDSPNYELPEEFSDLQGVLDDYGLLPHTLNRALALYQFHKLEGFCSDSYF DEWLEVLMQKDSPNYELPEEFSDLQGVLDDYGLLPHTLNRALALYQFHKLEGFCSDSYF DEWLEVLMQKDSTATRNSLASVEKGREIINHIEKAIKW

CENE 0088 106352 108145

CF288 hypothetical protein
SYRKCNOMYTVSECTAGCHVIEAYCHLLEVERFDGYVRQGEVAYVNVDNTWLKAEVIEVAD
GEVKVOVFEDTOGGRCGALVTFSGHLLEAELGPCLLCGIFDGLONRLEVLAEDSSFLORG
KHVNAISDHNLWNYTPVASVGDTLRRGDLLGTVPEGRFTHKIMVPFSCFQEVTLTWVISE
GTYNAHTVVAKARDAQKECAFTMYQRWFIKQAFIEGEKIFAHKIMDVGLRILDTQIPVL
KGGTFCTFOFPFGAGKTVLQHHLSKYAAVDIVILCACGERAGEVVEVLQEFPHLIDPHTGK
SLMERTCIICNTSSMPVAARESSIVLGVTIAEYYROMSLDILLLADSTSRWAQALREISG
RLEEIPGEAFPAYLSSRIAAFYERGGAITTKDGSEGSLTICGAVSPAGGNFEEPVTQST
LAVVGAFCGLSKARADARRYPSIDPLISWSKYLNOVGGILEEKVSGWGGAVKKAAOFLEK
GŞEIGKRMEVVGEEGVSMEDMEIYLKAELYDFCYLQQNAFDPVDCYCPFERQIELFSLIS
RIFDAKPYDSPDDARSFFLELOSKIKTLNGKLEFLSKEVILVBLIAFTMWOMA REFDAKFVFDSPDDARSFFLELQSKIKTLNGLKFLSEEYHESKEVIVRLLEKTMVQMA

108111 109466

CT289 hypothetical protein LDCWKKQWYKWRKDMQTIYTKITDIKGNLITVEAEGARLGELATITRSDGRSSYASVLRF LDCWKKOWYKWRDMOTIYHITDIKGNLITVEAEGARLGELATITRSDGRSSYASVLRF
DÜKKVTLQVFGGTSGLSTGDHVTFLGRPMEVTFGSSLGRRLNGIGKPIDNEGECFGEPI
EIATFIFNPUCRIVPRSMVRTNIPMIDVFNCLVKSQKIPIFSSSGEHHNALLMRIAAQTD
ADIGVIGGMGLTFVDYSFFVEESKKLGFADKCVMFIHKAVDAPVECVLVPDMALACAEKF
AVEEKKNVLVLLTDMTAFADALKEISITMDQIPANRGYPGSLYSDLALRYEKAVEIADGG
SITTITVTTMPSDDITHPVPDNTGYITEGGFYLRNNRIDPFGSLSRLKQLVIGKVTREDH
GDLANALIRLYADSRKATERMAMGFKLSNWDKKLLAFSELFETRLMSLEVNIPLEEALDI
GWKILAQSFTSEEVGIKAQLINKYWPKACLSK

CPn_0090 109439 110080 atpD-ATP Synthase Subunit D

AUDUTATE SIDINTE UVLAKSMSVQVKLTKNSFRLERQKLARLQTYLPTLKLKKALLQAEVQNAVKDAAECDKDYV QAYERIYAFAELFSIPLCTDCVEKSFEIQSIDNDFENIAGVEVPIVREVTLFPASYSLLG TPIMLDTMLSASKELVVKKWAMEVSKERLKILEEELRAVSIRVNLFEKKLIPETTKILKK IAVFLSDRSITDVGQVKMAKKKIELRKARGDECV

CPn_0091 110074 112053 atpI-ATP Synthase Subunit I

atpl-atp Synthase Subunit I Vernichaer Subunit I Vernikylfigenkadffsasrelgyvefiskkcfitteoghrfyeclkyfdhleaeys Vernikylfigenkadffsasrelgyvefiskkcfitteoghrfyeclkyfdhleaeys Lealefykdesysvedivsevitlakeikglletykalrkeivryrploafsseiaels Rktgislrffyrhkonedleedspnyfylstaynfdyylvlgyvdheorihoakeccedlfdgky Rheloydldinloreirnsporloldyayrevlrolchydneorihoakeccedlfdgky Favachvivdhikelgslcnryqiymervpvdpdetiptylenkgygmgedlygiydta Aysokdpstwyfaffylfpsmivdhoaydllfinksllffskrkrkkrskkhlsrhlkmtailggicuggttyfsffylfyshyvdyshlfidnilmelalfigvyhgpkaykeitney psikairdpkafllateigsagiesryvyydkfidnilmelalfigvyhlslighlrylrysskairdpkafllateigsagiesryvydkfidnilmelalfigvyhlslighlrylryssysfidvilfmysaylvydfylcytyslihylphypelggigygmfggiglavylami QRCWRGVeeilsvigysdylsylriyalglagammgatfnomgarlemllgsivillog Syniilsimggyihglrikhfiewyhysfdoggrplrplrkivcsedaeagihldnnsiv

112121

CFm_0992 L12121 112573 abpK-ATP Symbhase Subunir K EYLKCAHRVOM IDMSVVOPALVLGLAM IGGA IGCOMAGVACHAVMER IDEGHGKLIGMSA MPGCGG LYGF ILMLLMQAA IKNOTLSPVOG IA IGLOVGAALLVERVMQGKCC VEG IQAYA ROSSI YOKOYAA IG IVRSPSLFAVVFALLLL

112440 113015

GT.1093 1 LI2440 113915 CT303 hypothotical protein CKADVVSARPKIMEDIROYMORIMOREGEDNEFHCELEFERYYYSKLVFGETVELAAISV IGEGRESEPSESSFTEYVGPEY.SAAAQESIEOSCHDEVYGOOVVVTWSEPSRMRKCEPVT LYEWYYYSNGKVEKETYEVNQSAGYRVYCEKGEEYKEEQGIESTRVALCSGNOEIVSRRH

HUMMEV LULDGP

CPn_0094 113104 115971
va15-Va1y1 tRNA Syntherase
vyrvflsrdhkkfolrimttdpfekaynfodtepel//fweknomfkaeassdkppysvim
pppnvtgvlmdhalmyttdpfekaynfodtepel//fweknomfkaeassdkppysvim
pppnvtgvlmdhalmyttdpfekyrkmdsgfevorifdthastatoavverhloaseg
krrtdysredfikhiwawkeksekvvlsolpqloasdoddreftmeplanravkkafkt
Lfendy tyrgyvlvmdpvlgthaladdevoreekdywlyytryrmydsgesivvattree
professioner (yww.) and the color of the c V3VYYRSGAV (EPYLSKOMFVSV.UFAGALIZEV L.QDIKTE EKDFVKNY LJAVNHERDW CISRQLWWGHR FPVWYHKNDDERVLCYDGEG (PEEVAQDPDSWYQDPDVLDTWFSSGLWP LTCLGWPDENSPDLKKFYPTALLVTGHDILFFWVTRMVLLCSSMSGEKPFSEVFLHGLIF GKSYKRYNDFGEWSYISGKEKLAYDMGEALPDGVVAKWEKLSKSKGNYIDPLEMIATYGT GKSYKRYNDFGEWSYISGKEKLAYDMGEALPHGVVAKWEKLSKSKGNVIDPLEMIATYGF DAVRLITLGSCANREGGIDLDYRLFEEYKHFANKVWNGARFIFGHISDLGGKDLLAGIDED SLGLEDFYILDGFNQLIHQLEEAYATYAFDKVATLAYEFFRNDLCSTYIEIIKPTLFGKQ GNEASQSYKRTLLAVLLINVIGVLHPVAPFITESLFIRIQDTLGALPEGDGDAFTGHALR MLRSRACMEAPYPKAFDVKIPQDLRESFTLAQRLVYTIRNIRGEMQLDPRLHLKAFVVCS DTTEIQSCIPILQALGGLESIQLLDKEPEKGLYSFGVVDTIRLGIFVEEHLLKEKGRLE KERVRLERAVENLERLLGDESFCQKANPNLVVAKQEALKNNRIELQGILDKLASFA

118790 115956

CPI_UU95
pknD-S/T Protein Kinase
aCivcldreDQRsLeRYDIVRIIGKGGMGEVYLAYDPVCSRKVALKKIREDLAENPLLKR
RFLREARIAADLIHPGVVPVYTIYSEKDPVYYTMPYIEGYTLKTLLKSVWQKSSLSKELA
EKTSVGAFLSIFHKICCTIEYVHSRGILHRDLKPDNILLGLFSEAVILDWGAAVACGEE
DLLDIDVSKEEVLSSRMTIPGRIVGTPDYMAPERLIGHPASKSTDIYALGVVLYQMLTLS FPYRRKKGKKIVLDGQRIPSPQEVAPYREIPPFLSAVVMRMLAVDPQERYSSVTELKEDI ESHLKGSPKWTLTTALPPKKSSSWKLNEPILLSKYFPMLEVSPASWYSLAISNIESFSEM ESHLKGSPKWTLITALPPKKSSSWKLNEPILLSKYFPMLEVSPASWYSLAISNIESFSEM
RLEYTLSKKGLNEGFGILLPTSENALGGDFYQGYGFWLHIKKERTLSVSLVKNSLEIQRCS
ODLESDKETFLIALEOHNHSLSIFVDGTTWLHMYYLPSRSGRVAIIVRDMEDILEDIGI
FESSGSLRVSCLAVPDAFLAEKLYDRALVLYRRIAESFPGRKEGYEARFRAGITVLEKAS
TDNNEQEFALAIEFSKLHDGVAAPLEYLGKALVYORLQEYNEEIKSLLIALKRYSQHPE
IFRLKDHVVYRLHESFYKRDRLALVFMILVLEIAPOAITRGQEKILVWLKDKSRATLFC
LLDPTVLELRSSKMELFLSYWSGFIPHINSLFHRAWDQSDVRALIEIFYVACOLHKWOFL
SCSIDIFKESLEDQKATEEIVEFSFEDLGAFLFAIQSIFNKEDAEKIFVSNDQLSPILLV
YIFDLFANRALLESGGEAIFQALDLIRSKVPENFYHDYLRNHEIRAHLWCRNEKALSTIF
ENYTEKQLKDEQHELFVLYGGYLALIGGAEAAKQHFDVCREDRIFPASLLARNYNRLGLP
KDALSYOPBRILLBOKRIVHCHGUNHDFDILCOTMYHL LTEFEOL KDALSYQERRLLLRQKFLYFHCLGNHDERDLCQTMYHLLTEEFQL

CPn_0096 124347 118837

CT296 hypothetical protein

ETFLSILREFFMKSLPVYVSGIKVRNLKNVSIHFNSEEIVLLTGVSGSGKSSIAFDTLYA
AGRKRYISTLPTFFFATTITTLPREKVEEIHGLSPTIAIKQNHFSHYSHATVGSTTELFSH
LALLFTLEGQAARDHKEVLDLVSKEKVLSTIMELSEGVISILAPLLRKDIAAIHEYAQ
QGFTKVRCNGTIHPIYSFLTSGIPEDCSVDIVIDTLIKSENNIARLKVSLFTALEFGEGH
CSVLSDEELMTFSTKQOIDDVTYTPLTQOLFSPHALESRCSLCQSGIFISIDNFLLIGE
KLSIKENCCSFAGNCSSYLYHTIYQALADALFNNLETPWKDLSPEIQNIFLERGKNNLVLP
VRLFDQTLGKKNLTYKVWRGVLNDIGDKVRYTTKPSRYLSKGMSAHSCSLCKGTGLGDVA
SVATWEGKTFTEFQOMSLNNWHVFFSKVKSPSLSIQEILQGLKQRLSFLIDLGLGLYDYA
RALATLSGGEQERTAIAKHLGGELFGITYILDEPSIGHPODTEKLIGVIKKLRQGNTV
LLVEHEERMISLADRIIDIGPGAGIFGGEVLFMGKPEDFLMNSSSLTAKYLRQELTIPIP
ESREAPTSWLLLTEATIHNLKNLSIRLPLARLIGVTGVSGSGKSSLINNTLVPAIESFLK
QENPRNLHFEMGLIGRLTHITRDLPGRSQRSIPLTYIKAFDDIRELFASOPRSLRQGLTK
AHFSFNQPQGAGIGCGGIGTMTISDDDTPIPCSECQGKRYHSEVLEILYEGKNIADILDM
TAYEAEKFFISHPRIHEKIHALCSLRLDYLPLGRPLSTLSGGEIGRKLAHELLFASPKO QRAPANLHE MACIGENTITISDDEPT PCSECQERYHSEVLEILYEERNIADILDM
TAYEAEKFFISHPKHHEKHALCSLRLDYLPLGRPLSTLSGGEJQRLKLAHELLFASPK
TLYVLDEPTTGLHTHDIQALIEVLLSLTYLGHTVLVIEHNMHVVKVCDYVLELGPEGGDL
GYLLASCTPKDLIQLNTPTAKALAPYIEGSLDIPVVKSEPPSSPKSCDILIKDAYQNNL
KHIDLALPRNSLIATAGPGASGKHSLVYDILYASGNIAYAELFPPYIRQGLLKETPLPSY
GEVKGLSPVISVRKCSSNRSYHTIASALGLSNGLEKLFAILGEPFSPLTEEKLSKTTPQ
TIIDSLLKSYKDDYVTITSFIPLGSDLEIFLOEKQKEGFIKLYSEGNLYDLDERLPLNLI
EPAIVIQHTKVSPKNSSSLLSAISVAFSLSSEIWIYISQKKORKLSYSLGWKDKKGRLFY
EITHQLLSSDHPEGRCLTCGGRGEILKISLEEHKEKIAHYTPLEFFSLFFPKSYMKPVQK
LLKDENASQPLKLLTTKEFLNFCRGSSEFPGNNALLMEQLDTESDSPLIKPLLALTSCPA
CKGSGLDMYAMVVRINTSLLDIYQEDATFLESFLNTIGDDTRSIIQDLMNRLTFISKV
GLSYITLGQRQDTLSDGENYRLHLAKKISINLTNIVYLFEEPLSGLHPQDLPTIVQLKE
LVANNNTVIATDRSCSLIFHADHAIFLGPGSGPGGFLMDSDTEVCPSVDLHANNPOTEL
CPKAPLSISKANHTRGSDRTLKVNLSIHHIQNLKVSSPLHALVAIGOSVGSGKTSLLEG
FKKQAELLIAKGTTTFSDLVVIDSHPIASSORSDISTYFDIAPSLRAFYASLTQAKALNI
SSTMFSTNTKOGQCSDCQGLGYGWIDRAFYAJLEKRPCPTCSGFRIQPLAQEVLYEGKHFG
ELLHTPIETVALRFPFIKKIQKPLKALLDIGLGYLPIGQKLSSLSVSEKTALKTAYFLYQ
TPETPTLFLIDELFSSLDPIKKQHLPEKLRSLINSGHSVIYIDHDVKLLKSADYLIEIGP
GSGKQGGKLLFSGSPKDIYASKDSLLKKYICNEELDS

CPn_0097 124549 126006
pyk-Pytuvate Kinase
DSMITRTKIICTIGPATNSPEMLAKLLDAGMNVARLNFSHGSHETHCQAIGFLKELREQK
RVPLAIMLDTKGPEIRLGNIPQPISV3GGQKLRLVSSDIDGSAEGGVSLYPKGIFPFVPE RVPLAIMLDTKOPEIRLGNIPQPISVSGOQKLRLVSSDIDGSAEGGVSLYPKGIFPFVPE
GADVLIDDGYIHAVVVSSEADSLELEFMNSGLLKSHKSLSIRGVDVALPFMTEKDIADLK
FGVEQNMOVVAASFVRYGEDIETMRKCLADLONPKMPIIAKIENRLGVENFSKIAKLADG
IMIARGDLGIELSVVEVPNLQKMMAKVSRETGHFCVTATQMLESMIRNVLPTRAEVSDIA
NAIYDGSSAVMLSGETASGAHPVAAVVIMRSVILETEKNLSHDSFLKLDDSNSALQVSPY
LSAIGLAGIQIAERADAKALIVYTPGGGSPMFLSKYRPKFPIIAVTPSTSVYYRLALEWG
VYPMLTQESDRAVWRHQACIYGIEQGILGNYDRILVLSRGACMEETNNLTLTIVNDILTG
SEFPET

CPn_0098 127494 126071
No robust homolog present in Genebank/EMBL ab of 11/7/98
LVGKKFHQIKRTILEAPEYYLVGGIT/LVGRHTPEGFLYGIJ/KGEGFLAFYITGDYKKTAL
TNLALAFPEKTFDERYKIAROGLOHLITTLLELLAIDQD/YGINDKLITTVTTGRNFKGFS
SEEVIGIEDLEETFKNIGERGIGITEF GHOANWELPFLYITKNYHSIAFAKAITNORLSK
KITALPEVFKUKIVPFKNIGQGIEATINGKLIGITARUKULALLMGGYTYPLFGGPAFTTTS
PALLAZKTGFFVIAVNUSROAKGFFVITGAKLIGANKSILFMYTTVAILMGMMGFLEKGIA
GOLFOMMAIHKRWKRKIGNVIKKKYPZTHILVFYLOVGGHGFFFFLALAEGEGGTTTHLAL
GNADHLEELQFGFPEYGLIQLRNGGDITALERY ZPATEDLTNNLGHLYKHFRYTGGCAVY
GKRFLEKGLDHQAPLKNSILKITYTKIGKPAFPKNEKVEGFGETTYP

12/415 127565 137527

CPU_DUPP - 2778C - 1778C - 1777C - 1778C - 177

CPn_0100 129183 127882

CT011 hyporhetical protein
RTOKKTFILLDLETMIKFLSOLFIRHWPRKVVSLGFAIIIWILVQQSVTITRTLTNVPVR
RYOKKTFILLDLETMIKFLSOLFIRHWPRKVVSLGFAIIIWILVQQSVTITRTLTNVPVR
RYOKKTFILLDLETMIKFLSOLFIRHWPRKVVSLGFAIIIWILVQQSVTITRTLTNVPVR
RYDLUDAVOLGGFLNKKVSLTITOKKNTVQDLRPSNLEVVISAANHTESWIATID
KHNLVSVOHEINIRKDIHSVDANDIFVRLTQVVTEDILLTITKPIGSPPKGYEYLDVWPK
YLNQKVSGPKEYINALKEQGLELTFNLNKISFEELERNRIAQGSHDEIIFPIPKEMKKIL
PFENTFMDLNDPOADFRLEFLKRECIPLNLNFSFTFTTTOMPLEVSLDDVPP
HASBITHGTBIFANDAV AGE VANGANA FORD FORD OTWATITTICHEN
PENDETANGBITANGBITHDIALIFOT GPBB SE SEN NLCTARGEPUNLLEI-NBPANIC TKTKETTKLYKKEW

CPn_0101 129986 129141
ybbp family hypothetical protein
PSTLCNFSQYTTQGPSKTMPFDITYYTTPLLEIILIWVMLNYLLKFFWGTRAMDVVFGLL
AFLFLFVLADKLHLPIIRRLMLHVVNIAAIVVFIIFOPEIRLALSRIRFHGKKFFIDTQE
QFVEQLAASIYQLSERQIGALVVLENKDSFDEVLSFSSVKINATFSEELLETTFEPSSPL
HDGAVILRGDILAYARVVLPLAHDTTQLSRSMGTRHRAALGASQRSDALIITVSEENGSV
SLSRDGLLTRGVKIDRFKAVLRSILSPKEHKRKPLFSWIWKR

CPn_0102 130099 131466
cydh-Cytochrome Oxidase Subunit I
FYIQFWKFMDALILSRIQFGLFITFHYLFVPLSMGLSMMLVIMEGLYLVTKKQIYKQMTW FYIQFWKFMDALLISRIGFGLFITFHYLFVPLSMGLSMMLVIMEGLYLVTKKQTYKQMTW
FWVGIFALTFVLGVVTGIMQIFSFGSNWANFSEYTGNIFGTLLGSEGVFAFFLESGFLGI
LLFGRHKVSKKMHFFSTCMVALGAHMSAFWIICANSWOTFSGYEWVHKKGKLIPALTSF
WGVVFSPTTIDRFIHAVLGTWLSGVFLVISVSAYYLWKKRHHEFAKQGMKIGTICAVIVL
VLQLWSADVTARGVAKNQPAKLAAFEGIFKTEEYTPIWAFGYVVMEKERVIGLPIPGALS
FLVHRNIKTPVTGLDQIPRDEMPNVQAVFQLYHLMINLWGWWALTLISWSAYKGRWAL
KPFFLVILTFSVLLPEICNECGWCAAEMGRQPWVVQGLLKTKDAVSPIVQANQIVQSLVI
FSLVFIALLTLFITVLCKKIKHGPEEENDLTEFEVK

131465 132511

CydB-Cytochrome Oxidase Subunit II
NRGIFMELSLTSLLPLAWYVILGVAVFAYSFGDGFDLGLGAVYLKAKEDKERRILLNSIG NRGIFMELSLIFSLEPLAWY ILGVAVATSFGRAF DIGGRAVILGARTERIELENSIG
PWDONEWULVIIVGGLFAGFPACYATLLSIGTMPIWILVLLYIFRGCSLEFRSKSESVS
WKIFWDIIFICSGTAISFFLGTIVGNLIIGLPLSPDTSYASLSWILFFRPYAALCGAVVA
SAFAIHGSCFALMKTSDSLANRIAQCFPYILSSFLVFYVLFLGASLISIPKRFDAFFTYP
LILLIALTISCCVAAKTSVSKKRYGYAFIYSTINLLSLILISAATLTFPNILLSTVDPQY
SYTIYNSAVETKTLKSLLIIVLIGLPFIITYTCYIYRVFRGKTNFPSIY

CPME0104 133884 132676
CT017 hypothetical protein
EKSMMMLQISMLLLALGTAINSPAIYAADSQSVSFPEQLPSSFTGEIKGNHVRMRLAPHT
DTTIRFESKGDLVAVIGESKDYVISAPPGITGYVFRSFVLDNVVEGEQVNVRLEPSTS
APVLVRLSRGTQIQPASQEPHGKWLEVVLPSQCVFYVAKNFVANKGPIELYTQREGQKKI
AMDLINSALNFAHIELEKSLNEIDLEAIYKKINLVQSEEFKDVPGIQGLIQKALEEIQDA
YESKSLESSONTSIASSQCSTPKVSSSEVTTSLLSRHIRKQTALKTAPLTQGRENLEYSLF
RIWASMQQGNDHSEALTQEAFYRAEQKKKQVLAGVLEVYPHVVKNNPGDYLLKAQENTIA
FEYGTSINLEQWLGKRVTVECLPRPNNHFAFPAYYVVGIKEAS

CPn_0105 134883 134029
CT016 hypothetical protein
YVPFRKFSNONPMLLIYCKKKEIHLQWPQTAKIRFTPKIAMKVKINDQLICIPPFISARW
SQLAFIESQGEGMKDQGTLRLHLIDGKIISIPNLDQSIIDIAFQEHLLYLETSQSGKEDS
RDDDKLGVGVVLMNVLQQITKGNDIQVLPKNLISPLFSGTNPIEAILQHTPEHKDHPDAPT
DVALEKMADVIRVLSGNNATLLPREPPHCNCMHCQIGRVMNEEDTLAVSDKDLTFRTWDIM
OSGDKLYIVTNPLNPSDQFSVYLGPPIGCTCGEPNCEHIKAVLYT

CPn_0106 135073 136374

phöf-atpase

EKVRTOMKKTMVIDTSVFIYDPEALFSFENTRIIIPFPVIEELEAFGKFRDESAKNASRA
LSNTRLLLENAKTKVTDGVLLPSGSELRIEVAPLSNDDRRCKLLTLELLKIIAKREPMVF
VTKSLGRRVRAEALQIESRDYESKRFSFRSLYRGFRELQVSQEDIENFYKNGYLDLPLDV
VSSPNEYFFMSAGENHFALGRYYVSEGKIIALKAMDKSVWGIKPLNTEQRCALDLLLRDD
VKLOTLIGQAGSGKTILALAAAMHKVFDKETTYNKVLVSRSPIVPMGRDIGFLPGLKEKKLM
HWMQPIYDNMEVLFSINQMGNSSEALQALMDAKKLEMEALTYIRGRSLPKAFIIIDEAQN
LTHEIRIKTIISRAGKGTKIVLTGDPTQIDSLYFDENSNGLTYLVGKFHHLALYGHMFMTR
TEREFEIAAAAATTI. TERSELAAAAATIL

CPn_0107 137321 136392
CT058 hypothetical protein_1
KKSPPPVTPKEIPTOPKPPIPQRPEVSPTPTDHIVPGSIEASPILGKKPSPDSMVSPLSL
FHKMLLENWTPVEEPFPWPPAEKNQKIFAWALNQSKLIFVSTSGNIAQPRLVTDSMSMMI
WAANRTMSRDGAGTNQVLSAAVSVDSWGLSQRPLNPERGQTPLNEGECRAGMMRNADGS
NHTGKQGKPHYLAQLLGPKAVDHHNKSQAAFDRCKNAYLNCFSLAQTLGVTFLQIPLISS
GIYAPPENRKKRPNSEENKVRMRWIHAVKCALVAAMQEFGNEPGNTDRRMLIVLTDLKTPA ITDPKKKSHL

CPn 0108 137887 137303

CPD_9108
CTO18
KNLFHYKAILMSIFNEEVFIISHRHTPLGQTSTALRNTPLVNPLHRTNLQRIASYIPIFS
TFIGIKTLKGISSLQYSMVLMTGNFSSVCKTLPCPEIYEELPKVRKEAWLEIFGIKALYY
LVLGVIKIIKLIVRYLCPCCRPPEPREPQNPLTPTPLDMGQQIDAIFSTPTGFTGFKDFF LDDLLOEDKKKAPHL

CPn 0109 138646 141783

CPD_0109

11eS-Iso1eucy1-trnA Synthetase
RQKHTADEVGKNSFAKKEEQVLKFWKDNQ1FEKSLQNRQGKTLYSFYDGPPFATGLPHYG
RQKHTADEVGKNSFAKKEEQVLKFWKDNQ1FEKSLQNRQGKTLYSFYDGPPFATGLPHYG
HLLACTIKDVVCRYATMDGYYVPRRFGMDCHGVPVEYEVEKSLSLTARGAIEDFGIASFN
VVFFSTALOTPLSNFEASQNYKEVDDFSLVYKMPLQNDSASLLWYTTPWTLPSNMAIAV
OETLYVVRIQDKKCGEQWILSQNCVSRMF:NPEEFVILEDFSGKOLVGRTYEPFTFFOS
KREDGAFKVIAAGFYEESEGTGVVMMAPAFGEGDFLVCKENHVPLVCPQDAIKSFTEEIP
QYQGOYIKHADKEIIKFLKKEGRIFYHCTVKHRYPFCWRTDTPLIYKAVNCWFVAVEKIK
DKMLPANSCIHWVPEHIQESRFGKWLFGARDMAISRNRYWSTPIPIWKSADJELLVVGSI
RELEELTYTOTTOIHHHIFIDDLIYKDLNKEPHIPIPYPFOGVEDSGAMPYAGHHYPFENOK RELEGITYTYGTTOTHRIFETDDLNIVKDGREFHRIPYTTYTTPTWKSADJETLVVGST
RELEGITYTYGTTOTHRIFETDDLNIVKDGREFHRIPYTDCWEDSGAMPYAGHIYPFENOK
ETEGAFPADFTAGGLDQTRGWFYTLTVIGATLFDRPAFRNATUNGITLAGDGHKMSKRIN
NYFSHKYVLDTYGADALRIVTLHBVVVK NEDLRESDKGTEXVLKQTLLPLTITITLSFINTY
AGLYGFDEKGOTEPAYTCTDQWTLSNLYVVVKVRRGMSQYHLNFAVEPFTFTDDLTN
WYTHRIFMRAEDTPDRRAAFSTLYEVUTVTCKVTADFVPFLAEDTYQKLYLEKEPES
VHLGDFTQVEMDKTLDDLGKRMHDTRETVOLGHSLRKCHKLKVRQPLANFYT/GSKDRLS

LLKTFEGLIAEELNVKNVIT/EFAPRETYTT/KENFRMLGKKVGSKMKEVQKALSELPNN AIDKLIQEETWVLTIDDPEIALDGDDVVIGPHTDRGYTARGSALFSVILDCQLREPLIVE GIARELVNKINTMRRNQQLH/SDRIALRIKTTEAVHRAFLDYENYICEETLIIAYDFTQD SDFQGENMDINGHATQIEIT/SSIDS

143755 141827

CPD_0110 143755 141827

lepB-Signal Peptidade 1

SYPSIFMKOHYSLIKSPHILERSTYKLLKSKKLAHSPADKKQLQELLEQLEEAIFEHDQE

TA PLANCY AT THE PROTECTION OF A LIA TO BE INCOME.

KRYIKRCKKERPOJE LIF / JA TO BE TO BE A SIAA TO BE INDOMESTIONED INTOK

KRYIKRCKKERPOJE LIF / JA TO BE TO BE A SIAA TO BE INDOMESTIONED INTOK

KRYIKRCKKERPOJE LIF / JA TO BE TO BE A SIAA TO BE INDOMESTIONED INTOK

KRYIKRCKKERPOJE LIF / JA TO BE TO BE KREEPD VIGELENGIT VE PI SEDUT LIB TO BE A SIAA TO BE TO

143934 144761

CT021 hypothetical protein OLONRYPIMPNDSSTYFERILQKYLMKKQGKTLFLFLFLFSTAFSGLFASQTSSLRT OLOMNIFIAM TO THE STATE OF THE

CPn_0112 144743 145093
gatB-(Pet112) Glu tRNA Gln Amidotransferase (B Subunit)
DSDFGVVMMKNNTHPEYRQVLFVDSSTGYKFVCGSTYQSEKTEVFEGKEYPVCYVSVSSS
SHPFFTGSKKFVDAEGRVDKFLKRYSNVRQPAQQPQPEEDALPAAKGKKVVTKKKK

145329

CPn_0113 145329 146405
pfrA-Peptide Chain Releasing Factor (RF-1)
GFMKKVAEYLNRLAEVEIKISNPEIFSNSKEYSALSKEHSYLLELKNAYDKILNLEKVL
ADDKQALAIEKDPEMYVMLEEG INENKVELEKLNKILESLLVPPDPDDDLNVIMELRAGT
GGEEAALFVGDCVRMYHLYASSKGWKYEVLSASESDLKGYKEYVMGISGTGVKRLLQYEA GGLEARLY VAN HALSSAWAL IN LAUSASSAWAL GENERAL WAS SAUGHEN GTHRVQRVPETETQGRVHTSAITIAVLPEPSEEDTELLINEKDLKIDTFRASGAGGHVN VTDSAVRITHLPTGVVVTCQDERSQHRNKDKAMRILKARIRDAEMQKRHNEASAMRSAQV GSGDRSERIRTYNFSQNRVTDHRIGLTLYNLDKVMEGDLDPITTAMVSHAYHQLLEHGN

CPn_0114 146371 147261
hemK-A/G specific methylase
VMPTTSYSNMEIKKAIQEGTAYLDYYGVPLSDCEALYILMDLLEVSSRAKLFDLVGISET
MLMEYRKRLALRGQRCPTAYLNGAVSFLGLRLRVDSRVLIPRTETELLAEYIINYLLSHS
EIQTFYDICCGSGCLGLAIKKSCPHVEVVLSDVCPQAVAVANENAKSNGLDVKILLGDLS
APYTRPADAFVCNPPYLSFNEIIHIDPEVRCYEPWKALVCGSTGLEFYQRIAQELPKIVT
STGVGWLEIGSSQGESIKNIFSKHGIYGRLHQDLSGRDRIFFLEMDGRDPVSSGAYS

147279 148622

ffh-Signal Recognition Particle GTPase
MINSLSOKLSSIFSFLVSSRRINEENISESIREVRLALLDADVNYHVVKDFISKVKEKIL MINSLSQKLSSIFSFLVSSRRINEENISESIREVRLALLDADVNYHVVKDFISKVKEKIL
GEEIWKHVSPGQQFIRCHEELVAFLSDGREEFTIQKTPSIILLCGLQGAGKTTTAAKLA
DYVIKNKKAKKVLVVPCDLKRFAAVDQLKILVAQTKAEFYQSQENKPINVVVKALAYAKE
NGHDFVILDTAGRLNIDNELMEELTAIQKVSQANERLFVMNVAMSQDVLATVQAFDQSLD
LTGVILSMTDGDARAGAVFSIKHVLGKPIKFEGCGERIQDLRSFDPQSMAERILGMGDTI
NFVKEMREYISEEEDAELGKKLVTAAFTYEDYYKQMKAFRRMGPLRKLLGMMPGFNNAKP
SQKEIEDSEQMKRTEAIILSMTPEERKELVELDMSRMKRIASGCGLTLGDVNQFRKQMS
QSKKFFKGMSKGKMEQVRKKMSGGNQWR

CPn_0116 148592 148972 rs16-S16 R1bosomal Protein EKNVRRKSVALKIRLRQQGRRNHVVYRLVLADVESPRDGKYIELLGWYDPHSSINYQLKS ERIFYWLERGAQLSSKAEALVKQGAPGVYSALLSKQEARKLVVRKKRRAYRQRRSTQREE

CPn_0117 148983 150071
trmD-tRNA (guanine N-1)-Methyltransferase
TGMKIDILSLFPGYFDGPLQTSILGRAIKQRLLDVQLTNLRDFGLGKWKQVDDTPFSGGG
MLLMAEPVTSAIRSVRKENSKVIYLSPGGALLTAEKSRELAAASHLILLCGHYEGIDERA
IESEVDEEISIGDYVLTNGGIAALVI.IDAVSRFIPGVLGNQESAERDSLENGLLEGPQYT
RPREFEGKEVPEVLLQGDHKAISOWRLEQSERRTYERRPDLYLMYLYKRSIDHKFDEETT
TNRDHFKCDKISVVLEVNKLKRAKNFYCKVFGLDAMSCENKFCLPHEGKTIFMLREVQAE KKNIVTLSLSLDCACEEDFCYLLRRWELFGGKLLEKQADEHAVWALAQDLDGHAWIFSWH

CPn_0118 150075 150464
rll9-Ll9 Ribosomal Protein
KKENRWYIMVNLLKELEQEQCRNDLPEFHVGDTIRLATKISEGGKERVQVFQGTVMARR
GGGSGETVSLHRVAYGEGMEKSFLLNSPRIVSIEIVKRGKVARARLYYLRGKTGKAAKVK

CPn_0119 150520 151164
rnhB-Ribonuclease HII
LMMTSISELQRFLSMIAFEKELVSEDFSVVAGIDEAGRGPLAGPVVASACILPKGKVFPG
VNDEKKLSPKORAQVRDALNODPEVCFGIGVISVERIDQVNILEATKEAMLQAISSLPIS
PDILLVDGLYLPHDIPCKKIIQGDAKSASIAAASILAKEHRDDLMLQLHRLYPEYGFDRH
KGYGTSLHVEAIRRYGPSR\HRKSFSPIKOMCAIV

78

CPn_0120 IS1125 151778
gmk-GMP_Kinase
EELFONKANVCYCMNKILLOUPFOPDHOKCOPKLETIOAPAGVGKTTLVRMLEQEFSSAF
AETICVTRKPREGEVIONCYHPVOHEEFORLILDROALLEMVFLEGECYGTSMLEIERIW
CLOPHAVAVIDIQGALFIS:RMPSVOIFTAPPGGEELERRLASRGGEEGGORKERLEHSL
IDLAAANQFDYVIINDDLWGAYRVLKOIFTAREHRHIL

MENTALA MITTAN TANDAR MENTALAN SANTAN MENTALAN M UFCACYCOKLLALISYPIIPESAVAIWEMIJPKSLENCNLDTMYARDLWKEEILDVINEE FHLKSPRLLFTTVE

CPn_0123 155975 153774
recD-Exodeoxyribonuclease V (Alpha Subunit)
nSMEKICGYLEQILVENKDSGDITAYIKIPNKTTPILIKGKLPOPLELGSPIQIYGVWSH
SPSNTKYPQIHSYDSPLLYEYRGVFHYLTSKLIKGIGPKIAEKIIEKFQEKTCYVLDITP SPSNTKYFQIHSYDSPLLYEYRGVFHYLTSKLIKGIGFKIAEKIIEKFQEKTCYVLDITP
ERLSEVSGISETRCVSICKQLCEQKILRKTLLFLQEYNIPIHYGVRIFKKYQEKSIEKIC
EDPFLLAREMEGIGFKTADFIAMKLGVPRNSESRLCAGIQHSLEELQEEGHTCYPIELLI
DVVAKLLNQDVFDTPITLEEIDTQILNMQKRKLLHIODISGTLHWATRYLHLAEKTIVSD
LKRILFSSRRIRSIDCEKAIAWVEENLSIDLAEQQREAIKACFSEKLLIITGGPGTGKST
ITQAILK FEQVTHKIILAAPTGKAAKRMTEITGKHSVTIHALLQYBFTKKSFRKNHDNP
IDCDLIIVDESGMMDTHLLHHFLKALPDYTTLVFIGDIHQLPSVGPGNILKDLITSNKMT
VIRLNKIFRQVHDSGIVTNAHRVNEGELPILYSETGRADFLFFQKDDDEEALNHIIHLVT KFVPOKYHIYPODIOVLAPMKKOTLGIYMLNKALKHALNPKKANLHGRPOSYAVGDKVMO IRNNYNKEVFNGDIGYVSTINFEDKAVVVRMEGKHVGYSFSELDDLVLAYATSVHKYQGS ESPCIIIPIHTSHFMMLYRNLLYTAITRGKKLVILVGTKKAIAIATRNNRVQHRCTGLAE VLKELDTKKNYADL

CPn_0124 156575 158068

No robust homolog present in Genebank/EMBL as of 11/7/98
IRSKORTVAITLLVIGILLIASGIIFLAVAIPGLSSAVALGLGCGMTALGTVLLITGLVL
LIRSEKLALEQVEIKQARTRVNNELDQLSQYVFYTENVLDNLKRWSYRDLGFVRQAQEEV
TNLEQDIEEIFLTLRDIRNALDNEEFFMTHAKQCLAQVGESLFQDASIDEFINLAHLSEI
RQHLDINDPRWSMITKKVKGTVVRFIYVSTMYKQIKSNFEKSDFGQLRKMLLNNYKTIEE
VLYQSFQKGYNRAALLSEKTRIIHTSSLLHWEKDEDKHLNIKNECASRLENFKKFRTLFL
GLSEEDVIDFTGASGMCSKLPRKEVPLDGGKKKLPFKRTFADEQVGDWDRTTSLEHMTP
QEEDPLDRLMDQVEQEATSVLKDQDRYWKEIETSEAKFRSLPREDDFEKQSQIDSYIRDL
DDHLSVWANQLSAAEDALIEVTDVQEHGNREMLKNIQQGLELIEDAVKATLPRVDFIQEL
LEKEELPLVAARMSLENS LEKEELPLVAARMSLENS

CPF-0125 158072 158605

No. Fobust homolog present in Genebank/EMBL as of 11/7/98
KISSCAEIMSEVKPLFLKNDSFDLATQRFONLINMLQEQAEIYNEYEEKNARVONEIKEQ
KOEKKRIEDFEARGLGVLKEELASLTRDFHDKAKAETSMLIECPCIGFYYSIHQEEQRQ
ROEKLQKMAERYRDCKQVLEAVQVEQKDMISSRVVVDDSYFEEEKEEQKVDNRKKEQD

CPT-0126 158806 161085
No bust homolog present in Genebank/EMBL as of 11/7/98
LLYESYYCMGLFFFSGAISSCGLLVSLGVGLGLSVLGVLLLLLAGLLLFKIQSMLREVPK LLYSSYYCMGLFFFSGAISSCGLLVSLGYGLGLSVLGYLLLLLAGLLLFKIQSMLREVPK
APDLIDLEDASERLRVKASRSLASLPKEISQLESYIRSAANDLNTIKTWPHKDQRLVETV
SREERLAAAQNYMISELCEISEILEEEHHLILAQESLEWIGKSLFSTFLDMESFINLS
HLSEVRPYLAVNDPRLLEITEESWEVVSHFINVTSAFKKAQILFKNNEHSRMKKKLESVQ
ELLETFIYKSLKRSYRELGCLSEKMRIHDNPLFPWVQDQQKYAHAKNEFGEIARCLEEF
EKTFFWLDEECAISYMCXWPFINESIQNKKSRVDRDYISTKKIALKDRARTYAKVLLEEN
PTTEGKIDLQDAQRAFERQSQEFYTLEHTETKVRLEALQQCFSDLREATNVRQVRFTNSE
NANDLKESFEKT DKERVRYQKEQRLYWETIDRNEQELREEIGESLRLONRRKGYRAGYDA
GREWELLBRWKKNLEDWAHLFDATMDFEHFFYGKSFICSTABRIFTYE FEELMWADDIJAD REMELLRÖWKKNILRDVEAHLEDATMDFEHEVSKSELCSVRARLEVILEELIMMSFKVAD
IESELLSYSERCILPIRENLERAYLOYNKCSEILSKAKFFFPEDEOLLVSEANLREVGAQL
KQYGÖKKQERAQKFAIFEKHIQEQKSLIKEQVRSFDLAGVOFIKSELLSIACNLYIKAVV
KEŞIPVDVPCMQLYYSYYEDNEAVVRNRLIMMTERYQNFKRSLNSIQFNGDVLLRDPVYQ
PEĞHETRLKERELQETTLSCKKLKVAQDRLSELESRLSR

CPRLD127 162152 161130

yttffcationic Amino Acid Transporter
ESF#FPSANGESRTRNVPLGIFHGLVACLYWGIVFVIPNFLGSFGDLDIVLTRYTIFGIF
SLIAGAIKNPSVIKKTELYIWRKSLLWTLLINPVYYFGITLGIRYVGSAITVVIASLAPT
AVLVFSNTKOKELPYSLLFAISSVIITGVILTHLSALNLPTAASPLYSILGVIAVILSTS
LWVIYVIRNQSLLEKHPNLTPDTWSYLIGISALIICLPMIIILDLGGITHVTHNLISHTP
GSERLLFLLLCSAMGIFSSAKALIAWNKASLNLSPALLGAILIFEPIFGLVLTYLYSQSL
PSLQEGIGIFLMLGGSLLCLVLFGRKVQKSLENSQVSSSNE

CPn 0128 162262 163053

CPH_UIZ8 16262 163053
bpll-Biotin Protein Ligase
EDRGRMLRNQVLVYCSEGVSPYYLRHTIRFLKYYSTQEGAFDILRVDGNFLIKNPFWEET
TRLLVFPGGADRPYHRVLHGLGTARIFQYVSEGGNFLGICAGAYFGSKMIYFYEPEGAPL
QGARDLGFFPCTAKGPAYRGNFSYVSPSGVRVSPQLFSDFGLGYAMFNGGCFFEGSBGYP
GVNIESRYDDLPGKPASIVSRUSKGLAVLSGPHIEYLPHYCRMVKENVQKTREFLQRER
TTLDRYCQNLVQRLRQPAFSKADC

163747 163064 similarity to CT036

similarity to CTO36
DEQYILSHIHMDPRIFVTSEPLQKTYQKLQEKHVNNLGIASQVSLTDLQNKTQYENNLIE
TTTNEITYYFPVVHNPDILASEWDPISNQLYLIFKKFFIHYHNLFSTALERNQILLIDSL
NTGSSNPIARQMELLAFLCVFEQLDYNEDEYTIEPRDYFNRFVYKNSQTAPQIQSFGLLH
GYEEMSYASNNIRNVLTHSIVLCSPILYQLITEFDTTKIHADDFDCLI

163751

ONIQUES HOMOTO DESERT IN GENEVARIA SOL 11/7/98

DEMVKCOSTHENKKPAQLEPESKFAAITKESLATLOLFLOTAACILIALSGLEPHTLLI
FALOLISTIVETTSISLLIOTOCOKSVOKDEQKPKOTEPKETPSLDPWULNPLKNKIQSS ETULLDPTS INLKNEUTFPSFEEWKK I FLK DPDFL I KSALANWK I LE

164441 165580

CITE_D131
No robust homotos prisent in Genebank/PMBL ac of 11/7/98
PSSSLKKERFSLSSLATI GIPPPTSAYVPDST CEUFLEMENAMERSEVAN POWITKTSVAQ
EVPRKHISKS (QVILETISMILETISMILETERSETERSYLTYPVT) TALIMILA GILVIELL RES
ESSMOMIA MISTER YALHQHENGEFLOVREVSQ GILLES BY IKVRALWESSD (FEDESAA
VILLESWIPPESSVOVE ALLESI QERECKY LOFST, PET, IR TERVISLIVELSAFTIDDENEQ
GVMT-MINISER DET INKARGIG TODIKHE IMSTGERTSVEDISMI TOVSOAMESVYRY
ROBLIESESTERSHILLS SERVINGER EROPDIAPEDELECTREDICTEGOVVHCLA TENPKOLADEDFEFACKNVEWDEFTSACEK

ALLKNPOGISIKDLKOFLVP

165584 166561 CPn 0132

CPn_0132 165584 166551
No robust homolog present in Genebank/EMBL as of 11/7/98
SMIEFAFVPHTSVTADRIEDRMACRMNKLSTLAITSLCVLISSVCIMICILCISGTVGTY
AFVVGIIFSVLALVACVFFLYFFYFSSEEFKCASSQEFRFLPIPAVVSALRSYEYISQDA
INDVIKDTMQLSTLSSLDPEAFFLEFFYFNSLIVNHSMKEADRLSREAFLILLGEITWK
DCETKILPWLKDPNITPDDFWKLLKDHFDLKDFKKRIATWIRKAYPEIRLPKKHCLDKSI
K MORTH LINDSDAVIOLE LINGUOYEGEFPANTIGLGSEVPMATGLBKVPMDLTWFMF
MEIMICILL NORD AVA

1 N.T. WEAREST OF CASE ONE

CPn_0133 167349 166564
CHLPS hypothetical protein
NSSAYMFKLLKNLFLIGCCIVGYFWMRKESIVEQWLSNRLHTQVTVGRVSIRTSGIKIRH
ICIHNPLASERFPYAAEIEYADVFFSSISMLLTKQLEISELIIHGANFTIFFYDSHGTKT
WMSLWWKNPHPQKETPSNLMIDRAPVLIRRCLFLNTRLYGLRANKHDIPHLSVPSLEFHS
HTSSAKELPKLSEALPSLLYLALEESLYHLNLPGDIIKPLSQQAHKHFYSSYPQFQDRLN
DINTPGTPTEEIIGFIRGLFFH

169131 167467

gtoel-HSP-60 FADYRKLRRTTMAAKNIKYNEEARKKIHKGVKTLAEAVKVTLGPKGRHVVIDKSFGSPQV TKDGVTVAKEIELEDKHENMGAQMVKEVASKTADKAGDGTTTATVLAEAIYSEGLRNVTA GANPMDLKRGIDKAVKVVVDELKKISKPVQHHKEIAQVATISANNDSEIGNLIAEAMEKV GANPHDLERGIDRAVKVVVDELKRISZFVVDHREFIAQVATTSANNDSEIGNLIAEAMEAV
KKNSITVEEAKGFETVLDVVEGMNFNRGYLSSYFSTNPETYGECVLEDALILIYDKKISG
IKDFLPVLQQVAESGRPLLIIAEEIEGEALATLVVNRLRAGFRVCAVKAPGFGDRRKAML
EDIAILTGGQLVSEELGMKLENTTLAMLGKAKKVIVTKEDTTIVEGLGNKPDIQARCDNI
KKQIEDSTSDYNKEKLQERLAKLSGGVAVIRVGAATEIEMKEKKDRVDDQHATIAAVEE
GILPGGGTALVRCIPTLEAFLPMLANEDEAIGTRIILKALTAPLKQIASNAGKEGAIICQ QVLARSANEGYDALRDAYTDMIDAGILDPTKVTRSALESAASIAGLLLTTEALIADIPEE KSSSAPAMPSAGMDY

169448 169143

groES-10KDa Chaperonin
MSDQATTLRIKPLGDRILVKREEEEATARGGIILPDTAKKKQDRAEVLVLGTGKRTDDGT
LLPFEVQVGDIILMDKYAGQEITIDDEEYVILQSSEIMAVLK

171419

PepF-0ligopept:dase
KGVPSLMTTELKTEALPTRTQVDPKHCWDTTLMYANREEWKKDFDLCSSGKDRSPIWPEF
SPSHYQIDNPESLLELLSKKFSVERKLDQLYIYAHLIHDQDITNPEGESDYQSIVYLYTL SPSNIQIDNPESLLELLISAN SVERNALDUI I IAHLI HUQUI INFEESDIQI VI III FSQEISVIQI VI III FSQEISVI PSQEISVI P LSNMWURTENRIKRSGATSSGATURFTTILLNTTINLLIUVSVTAHEAGRARISTFSREA OPYHDAQVPLFLAEIASTFNEMLLMEALSKSDQSKEDKUVIITKTLDTIFATLFRQTFFA AFEVEIHSAAEQOTPLTEEFLSATYGNLQKEFYGGVVTSDSLSALEWARIPHFYYNFYVY QYATGIIAALSFAEKILTQEPGALELYLKFLKSGRSDFPLNILKKSGLDMTTSAPLDKAF AFITKKIDLLSSLLSED

172263 171502

YbgT-ACR family
YbgT-ACR family
VCSMNVADLLSHLETLLSSKIFQDYGPMGLQVGDPQTPVKKIAVAVTADLETIKQAVAAE
ANVLIVHHGIFMKGMPYPITGMTHKRIQLLIEHNIQLIAVHLPLDAHPPLGNNMRVALDL
NWHDLKPFGSSLPYLGVQGSFSPIDIDSFIDLLSQYYQAPLKGSALGGPSRVSSAALISG GAYRELSSAATSQVDCFITGNFDEPAWSTALESNINFLAFGHTATEKVGPKSLAEHLKSE FPISTTFIDTANPF

CPn_0138 174094 172700

"hemL-Glutamate-1-semialdehyde-2, 1-aminomutase"
TNSRLFLAIKDOLLQNMWRLTKRNSMLINCSNQKHTVTFEEACQVFPGGVNSPVRACRSVG
VTPPIVSSAQGDIFLDTHGREFIDFCCGWGALIHGHSHPKIVKAIQKTALKGTSYGLTSE
EELFATMLLSSLKLKEHKIRFVSSGTEATMTAVRLARGITNRSIIIKFIGGYHGHADTL
LGGISTTEETIDNLTSLIHTPSPHSLLISLPYNNSQILHHVMEALGPQVAGIIFEPICAN
MGIVLPKAEFLDDIIELCKRFGSLSIMDEVVTGFRVAFQGAQDIFNLSPDITIYGKILGG
GLPAAALVGHRSILDHLMPEGTIFQAGTMSGNFLAMATGHAAIGLCQSEGFYDHLSQLEA
LFYSPIEEEIRSGGFPVSLVHQGTMFSLFFTESAPTNFDEAKNSDVEKFQTFYSEVFDNG
VYLSPSPIEDANFISSAHTEFNITYAQNIIDELIKTEDSAQDFF VYLSPSPLEANFISSAHTEENLTYAQNIIIDSLIKIFDSSAQRFF

CPn 0139 174686 174093

yggE
SPTRNKLRDIMKIPYARLEKGSLLVASPDINQGVFARSVILLCEHSLNGSFGLILNKTLG
FEISDDIFTFEKVSNHNIRFCMGGPLQANQMMLLHSCSEIPEQTLEICPSVYLGGDLPFL
QEIASSESGPEINLCFGYSGWQAGQLEKEFLSNDWFLAPCNKDVVFYSEPEDLWALVLKD LGGKYASLSTVPDNLLLN

CPn 0140 175140 174673

yddi. Prsnogkifcmslekelleetplullnfyklusfcnyagmilgteekkfaiyghusmgqa FQGADTEGHSPQRPFAHDLLNFVFSGFDIQVLRVVINDYKDNVFYTRLFLEQKDREFLYV VDVDAPPSDSIPLALTHKIPILCVKSVFDAVVPYEE

175817 175110

CPT_0141 175817 175110
rp1A-Riddse-5-P Isomerase A
HSSSAVEYDLHLHEKKCLAHEAATQVTSGMILGLGSGSTAKEFIFALAHRIQTESLAVHA
IABSQNSYALAKQLAIPLLNPEKFSSLDLTVDGADEVDPQLRMIKGGGGAIFREKILLRA
AKRSILIVDESKLUPVUGKFRVPLEISPFGRSAIIEEIRHLGYEGEWRLQDTGDLFITDS
SNYIYLIFSPNSYPNPEKDLLKLIQIHGVIEVGFVIEKVEVWSSNSQGLISKKYSV

176121 175816

PRIGHTAY NO TOBUST DOMESTOR PERSON TO SERVICE DESCRIPTION OF TOBUST DOMESTOR PERSON OF THE TYPE OF THE TOBUST DESCRIPTION OF THE TYPE OF TYPE OF THE TYPE OF THE TYPE OF TYPE OF THE TYPE OF TYPE

GARKA INDPADAGGALOPDA, ENCORADNIKA CMAGIDEKOPODPIOÁTPINURA IND NEJIMILACPKENEJEDNACHEDANACHENNAHHEVENA ELEMENALHEN INDIATOPPIKE ORVANIPELEN PERENACHENNAHHEVEN MELEN METERAKALIMOFITALE EGINA NEJIMILACPKEN PELENALHAN PENGUNAHHEVEN MELEN METERAKALIMOFITALE GOTA NEJIMILACH PROPERTATION TO TO TO TO THE TO T

CPN_0144 177942 180560
c1pB-c1p Professe ATPase
KVLGVNFMEKFSDAVSEALEKAFELAKSSKHTY/TENHLLIALLENTESLFYLVIKDIHG
NEMLINTAVKDALSREPTVVEGEVDPKPSPGLOTLLRDAKQEAKTLGDEYISGDHLLLAF
ACHBEGFERBWE JOHN PROFILENDE TO A TRANSPORT OF THE SEASON OF THE SEASO SKALLKGDIKPDTSIELTMAKEVLVFKKVETPS

180717 182369

CMI_U145

T30/1/ 18259

CT114 hypothetical protein
NCAASFIWLNKSSHRNLRSPMFKSFIVRYMFVCGLVSFLLPIPDLECANNVTKTYDKKAS
VISRDLKLQEDCQKFWNLDPYKLESLCAYQVLYHDDYSSKRIRELFPQIQKDEVPIFATM
LITLGKVNPGFSPEEISLIOKLSYPGLSLASLRGSTEIDPNTDLARALVVSEFSGDLGKN
RADYYSNCLDILALRIHAERQRYLDQSPCVPGTSEFHKATIEAINTILFYEEAVRYPSKK
EMFSDEFSFLSSVTDRKFGVCLGVSSLUFSLSQRLDLPLEAVTPPGHIYLRYQGEVNIE
TTAGGRHLDTASYCDCLDLEBIOVRTPEFMIGLTFMNGGSPALAKKVYGAFFAVKAOF EMF 3DEF 3F LSV TOKRYOLGVSSTE SLSQXIDE BLEY TYTAGGRILPTASYCOCLOLEDLQVRT PEEMIGLT FYMOGSFALOKKYKEAEEAYKKAQE YLGDELQELLGFVQ ILGGKKEGKSLIGKSPRASOKGSVAYDYLKGRINI PYLALLFSY YLGDELQELLGFVQ ILGGKKEGKSLIGKSPRASOKGSVAYDYLKGRINI PYLALLFSY PGSNYEEIASYEEELKKAMKSSMPCCEGQRRLASVAFHLGKTAEAVALLEKCVEDI PNDL SLHLRLCKILCDRIEYTKALKYFI IAERLMEDQGFLKKDNRSFALFYEVKKI ISKVAPQK

CPn_0146 182595 183095
NO robust homolog present in Genebank/EMBL as of 11/7/98
IIVGISMSSSEVVFQTVHGLGFGGLSKSVVPFKKSLSDAPRVVCSILVLTLGLGALVCG
IAITCWCVPGVILWGGICAIVLGAISLALSLFWLWGLFSNCCGSKRVLPGEGLLRDKLLD
GGFSRAAPSGMGLPGDGSPRASTPSCLEELQAEIQAVTQAIDQMSDD

TEPPLEGGVAGEAGGGGSPLTQLDLNSGAGS

CEME 148 183822 185702
pkml-s/t Protein Kinase
gkwwnyssmesekdigakfiadyrilyrkgoslwsedllaehrfikkrylirillpdigs
sqermeafhdvyklakinhegilsienvsesegreflytoeodipilsltoykksiprk
LTEEDIVDIVSQLASLLDYHSEGLAGEEWNLDSVYIHILNGYPKVILPDLGFASLIKER
IIDDEFISDEENRESKIKERVLLHTSEGRGGREDTYAFGATTYYLLFGFLPOGIFFMPSKV
FSBFIYDMDFLISSCLSCFMEERAKELFPLIRKKTLGEELDNVYTNCIESSLREVPDPLE
SQNLPQAVLKYGETRVSHOOKESAEHLEFTLVEACSIDEAMDTAIESESSSGVEEGGYS
LALOSLLVREPVVSRYVEAEKEEPRPOPILTEMVLIEGGEFSRGSVEGGRDELPVHKVIL
HSEELDWADTTAFGETRYLFCGSSODKYWNFILDFENDSTORPRSGFLYFEDVAKHDUV HSFELDVHPVTNEQFTRYLECCGSEQDKYTNELIRLRDSFTQRRSGRLVIEPGYKKHPVV
GVTNYGASGYAEWIGKRLPTEAEWEIAASGGVAALRYPCGEEIEKSRANFFTADTTTTMS
YEMEPYGLYDMAGNUYEWCODWYGYDFYEISAQEPESPQGPAQGVYRVLRGGCWKSLKDD
LRCAHRHRNNPGAVNSTYGFRCAKNIN

CPn_0149 dnlJ-DNA Ligase 185706 187700

CMIT-DNA LIGASE
EREMKEENSQAHYLALCRELEDHDYSYYVLHRPRISDYEYDMKLRKLLEIERSHPEWKVL
WSFSTRLGDRPSGTFSVVSHKEPMLSIANSYSKEELSEFFSRVEKSLGTSPRYTVELKID
GIAMAIRYEDRVLVQALSRCNGKQGEDITSNIRTIRSLPLRLPEDAPEFIEVRGEVFFSY
STFGTINEKQQQLEKTIFANPRNAAGGTLKLLSPQEVAKRKLEISIYNLIAPGENDDSHYE
NLQRCLEWGFPVSGKPRLCSTPEEVISVLKTIETERASLPMEIDGAVIKVDSLASQRVLG NLQRCLEWGFPVSGKPRLCSTPEEVISVLKTIETERASLPMEIDGAVIKVDSLASQRVLG
ATGKHYRWALAYKYAPEEAETLLEDILVQVGRTGVLTPVAKLTPVLLSGSLVSRASLYNE
DEIHRKDIRIGDTVCVAKGGEVIPKVVRVCREKRPEGSEVWNMPEFCPVCHSHVVREEDR
VSVRCVNPECVAGAIEKIRFFVGRGALNIDHLGVKVITKLFELGLVHTCADLFQLTTEDL
MQIFGIRERSARNILESIEQAKHVDLDRFLVALGIPLIGIGVATVLAGHFETLDRVISAT
FEELLISLEGIGEKVAHAIAEYFSDSTHLNEIKKMQDLGVCISPYHKSGSTCFGKAFVITG
TLECMSRLDAETAIRNCGGKVGSSVSKQTDYVVMGNNPGSKLEKARKLGVSILDQEAFTN
LIHLE

187759 192444

CPn_0150 187759 192444

CT147 hypothetical protein
IIYYKFFYSYNCPYFISFFVLLGVMMASSSNNSTKQLGIPSWVNPNVQWNRASQVGDQEA
NSLTPEAGYGSKNFSDRKHFLEVLDVSLEEMENNDLKKYSRYKTIILIATLVTVAITCIV
PISMVFGIPMWVPCLILFGAGLSSAFLSHRLQSKCKEIHLRYRAYQIYRQQLLSQYPDLR
KSTLYKYSITHVKPKKGFVGKLVENLRPDLHKNKDDGAAADSRLDFAGYGYKYYQTDAL
LGVSGVNSVEWQRLASLIMSVKNDILNDVGSREPIDMAQRSALVVSGKDIGGEIQPGGIL
DISRDILAICGYGNNVGVEAKKAIDQYKKWYLNSSTFIAWNPQLPAIAQSYLLEQQRHLD
VAAKIFQDLSALTTAHOTQQALEDLDSLLCYYDQLIESKGVGEKIIASIHQKHLDLAMQD
DCDGEHLKKWSNLYHVFSITIKEFTEGKLEQNEVVSRIQRLRGKLEKSKCSILGNCRTNA
EYATKSEKKLADVLLQIGDREPFLTOMHKAIATGKAIQGKVEGVISQHPEKQIMMLRCSI
EPLEGMLRREDWGAILGKNEDEVLALKSTMEAQLQGFKDLOTWEGKYGEFKKNKLSKVL
V7DFTKSYSNLLARLEVLHAESSTDDLVLHVDRMSEDLKKTIEEIDGNLFQVTPEELSL
APEYQLANNELPILVQEGNRLQEAIGGEGVSQGLMLHIGLLNRDEKINKNIESSRRNLVA
AAPQAR:DARNIDJGGLAPLIQRNAASLDNILQMMYLFNGSIRNIHALDTETLVATSSNM
EGAMHTFDWNIYTNLLDVLETQSKFAPAFMENFDLFGALPESVQDAVAEDVGGTHRHHQ
VLERGADDLKNM ISQLOKSINKWGMAKAIVLGIVAY/LFCVLSAIFTGONILGLLLISCVG
VLERGADDLKNM ISQLOKSINKWGMAKAIVLGIVAY/LFCVLSAIFTGONILGLLILSCVG
ULDTOVCHLIFDRISKSKEPEKQVLETAGSLIPATYILDGEFNNKDLARLAKLQDNLNLE
GRYFTWARNIVSDLEGITTKEKGLKDLKDITKEFRRD DYNLINKRIKRFEKEGLIJGEAPVYRPT
TEGDITRSEGVKRAKRKFKENVLHILAKRGOLKDLELLELTYTTAGGSLCRHGWQAGFRERILLINF
KYGKLON IDRIJCKNNLETYVRIONFFRTLIQEKLGPLTYVQEIDVYKEAKELHELAAITYG
NICHREGERFTLAIREEHLGYDCOKEELJIRGENLYTTAGGSLGKRYDREKAHAAAMT
KKYGKLON IDRIJCKNNLETYVRIONFFRTLIQEKLGPLTYVQEIDVYKEAKELHELAAITYG
NICHREGERFTLAIREEMLKFUGLGYLTPEVRSIPFTTAGGSLGKRYTTEMBOLHAVEELUF QUQETY::PEDYAAVQAALAAYYRKHEGLIV:STYGLGAQEXQTEGRYTTLMRDLHAVEELV EMCVETYRLNR::XQTLHRVH::VLHSHLRDCDSSGNGTIDVVKKLFELLNNNCNNPNDPEC

QKYMQILLDAPVSLLYGAFKSFKNEFULNITTELNIANSTKAAEEEAKRYVEEKGRGFETY WEEAKQRLEAIAAELDDLRNGETLLEGEIPLANLKISIFSDLNLREKVSVEKAALEEEIQ GIQEQYAEMQGIEDLELKQKFEDLQKKLEALEERLLQIGRRIDSSVDKQKELLGLLGREE

134178 132625 CPn 0151

CPN_0151 194178 132625

mhpa-Monooxygenase
CYENLEHYPRASMADILVIGANPTGLILANMLIQHGISVKVIDHRASPEDPSFLDCRKLP
VILANINI TILBNOPRI ODE 10 NAMIL TO 132 MANYPOTLLEKEGO ATTO 122 MEDICALE TO 122 MANYPOTLEKEGO ATTO 122 MEDICALE TO 122 MANYPOTLEKEGO ATTO 122 MEDICALE TO 122 MANYPOTLEGO AND 122 MANYPOTLEGO ATTO 122 MANYPOTLAGO AND 122 MANY

CPn_0152 195274 194318
CT149 hypothetical protein
Likwrkvaflvsclfsvaigasaapvrvygfpojpedlvoiktevcpkgevclavtikcd
DhNLigVihlpntptpeggfptvvlfhgfrgtkfggltgayrkLgrkfaavgiatlrvDm
AgcGobsevaeevpietylrdaofiletvogehpdlmayrlgisgfslochiafelakiyn
PRDLNikalswapiadggillkelyenfskhgegdiisvgkdfgfgpppiivcsgdvdl
Liriodhytamslptkpyilhoggiddtlvsrtogtlfkntapgrmffisypntghnlat
Abbliphildivshedbti APDLDMILDQIVSHFQRTL

CPn_0153 195430 197892

1euS-Leucyl trna Synthetase
nmrydpnliekkwoofwkehrsfqanededkvkyyvldmfpypsgaglhvghligytatd
Ivarykrargfsvlhpmgwdsfglpaedyairtgthpkvttoknianfkkqlsamgfsyd
gerfeatspddyyhmygklflflydgglaymadmavnycfelgtvlsneevengfsiegg
ypverkmlrowilkitayardkllegldaldwednvkgloknwigksegalvtfhltoges
Leafttrldtlgvsflviapehpdldsivseedpodbytavvogslrksgedrissyktk
tgyffgnyakhpitgrllpwisdyvvlgygtgvvmgvpahderdrefaemfslpihevi
DDNgvcihsnyndfclnglsgoeardyvinylemrslgraktmyrlkrublfsgrorwgep
IPIIHFebothrpleledelpllpbepfcgoplakaodwytydettgrorger
Tytmpomagscwyylrfcdahnsolpwskekesympvdlyiggaehavlhllysrfwhr
Vfydaglvstpepfkklinoglvlassyripckgyvsiedveengthistcgeivery
gemskskingvdpovieeygadalrymymfsgpldknktmsnesvgcorrflnrydly
tssevodiedroglvlahklyfritehtekmslntipssfmffsklpvyskralsm
avrvlepiaphtiseelwvilgnppgidoampoidesylvagtvtfvvovnsklrgrlev
kreapkeevislsrsvakylenagirkeivypnklvnfvl AKEAPKEEVLSLSRSVVAKYLENAQIRKEIYVPNKLVNFVL

197874 199202

gseA-KDO Transferase GSEA-KDO Transferase
TSEFCPMMLRGYHRIFKCFYDVVLVCAFVIALPKLLYKMLVYGKYKKSLAVRFGLKKPHV
PGEGPLVWFHGASVGEVRLLLPVLEKFCEEFFGWRCLVTSCTELGVQVASQVFIPMGATV
SILPLDFSIIIKSVVAKLRPSLVVFSBGDCWLNFIEBAKRIGATTLVINGRISIDSSKRF
KFLKRLGKNYFSPVOGFLLQDEVQKQRFLSIGIPEHKLQVYGNIKTYVAAQTALHLERET
WRDRLRLPTDSKLVILGSMHRSDAGKWLPVVQKLIKEGVSYLWVPRHVEKTKDVEESLHR
LHIPYGLWSRGANFSYVPVVVDEIGLLKQLYVAGDLAFVGGTFDPKIGGHNLLEPLQCE
VPLIFGFHITSQSELAQRLLLSGAGLCLDEIEPIIDTVSFLLNNQEVREAYVQKGKVFVK
AETASFDRTWRALKSYIPLYKNS

CPn_0155 199697 199498
No robust homolog present in Genebank/EMBL as of 11/7/98
NSLSFGVPFLEKLKISLIPIEEMRHELFMKTHNSSSNGFSNQEKGIRTYFKSDLLGYEDL YFLRENINPN

CPn_0156 200147 199770
No robust homolog present in Genebank/EMBL as of 11/7/98
LGKQKLLARHMHNIVVLSEEPGRSAFLGRTAFFPNKYPIAQGGVGIPSTIGNLFTIWYC
FYFYRAATPQSDHPDGCGFILLERLKELGAGFFYCDLRESNTTGFTLFFEGSNKGVLKNH LFIRDE

CPn_0157 200753 200298
No robust homolog present in Genebank/EMBL as of 11/7/98
FSFVTYKEALNNIYOFSPCASPHWQASLMAQLNSYFCLGGETVTRIISLRPSGLILAKKE
KAVVSTAEKNILKILSFILFFLVLTALATRYLLYNKFNKDLDRAVFFIPTEITKAEELIIA
KNPALVKEAALTVSPLFYSLPKKYQLMKVETP

CPn_0158 201463 200894 No robust homolog present in Genebank/EMBL as of 11/7/98 PPKITLSINIDLLLEDLDTDSIPWPKLYLSEDFDFAYYPESKAIIDTVAKLEKNNPGEFF CLESKKILARYLLECLFKLETGLMFPTSTIDGGRESFLIEFSHETKKPTVWAFTYFYYYH SNGPKLEKDFKQAGCEVHNRLLNLGLKYRPQAGAQNDGRNGGPYGPIGFLIVWEENYGSV LKDHGFIKDN

CPn_0159 201811 201467
No robust homolog present in Genebank/EMBL as of 11/7/98
CCPGGETATRIFSMTPSGFSLATEEK/JQVSTAEKVIKILALIFFPIILIALATRYFLHRK
FDRKCFVIPQDTPKELELILAANPQLVEKAAREVHFGFFALPTKYQSMYIQTSKG

203794 202127

CPA_0160 203794 202127
pfkA-Fructose-6-P Phosphotransterase
TVELLSLNKSYFEIQRLRYRPEILTLLETIRSKHIQETSSPPSPPPELQKHIPNLCRIPE
VSIYTEOETSSKPLKIGVLLSGQQAP75HNVVIGLPDALRVFNPKTRLFGFIKGPLÖLTR
GLYKDLDISVIYDYVNMGGFDMLSSGPEKIKTEGOKKNILNTVKQLKLDGLLIIGGNISN
TDTAMLAEYFLAHNCKTSVIGVPKTIDGDLKNCWIETSLGFHTSCRTYSEMIGNLAKDAL

HSPFGALYSTSGKRSL

CPH_0161 204658 203798

(prodicted acyltransiciase family) LABURTGSKRUHVRLAGELTREG LAALPVOLLURIGDCEGELMDEGLENYKON FRE LEEYT HOLLH LOGERLA LEGUSLAGELALQTOLFERK LKALAVWALTEGGELMAAFAQENAPEV L

TM::QKGAITYAGMTLNPDFYTQFLKIDIVKELMPSARNLPPILYMGGEQDLLVSINHRTL FTEAFANQDKPITILTYPDVDHAFPFAESSALSDLTQWLKRELTSGE

205870 204803

No robust homolog present in Genebank/EMBL as of 11/7/98 NO TODUST NO TODOST OFFSENT IN GENERALLY BUSINESS OF TITTING PRIMARYS ISSURVET WIFE CHARKVEST AS OF THE TOTAL OF THE CHARKVEST OF THE CHARKVES PLDEDRGGGFEILEQLQELGVRFPICPSQGPDNPNFQGFQGIRIYWEDSYQPNKEV

CPn_0163 205831 206394
No robust homolog present in Genebank/EMBL as of 11/7/98
FEKAIVYCIKCKQIIKCISIIHTPTPATPLCTEGEIFPOPVDSAIQNDLERLLTVKKRPD
IIREYLRAGGSLVTTYPKECQRLKSPEQLRVLDDLVQSYPNHLHAIELDCGAIPQDLIGA TYIITFADFSTYILSLRSYQANSPSDDTWGIWFGSIDDPVQAVISFLKDHGFALPSTLAQ

CPn_0164 206444 206998
No robust homolog present in Genebank/EMBL as of 11/7/98
LCFKCIYIKIIFSFLKQLMTRSTIESSDSLCSRSFSQKLSVQTLKNLCESRLMKITSLVI
AFITLIVGGALIALAGGGVLSFPLGLILGSVLVLFSSIYLVSCCKFFTLKEMTMTCSVKS KINIWFEKQRNKDIEKALENPOLFGENKRNVGNRSARNQLEMILHETDGIILKRYMKGAK MYFYL

CPn_0165 206983 207582
No robust homolog present in Genebank/EMBL as of 11/7/98
NVLLFMNWVPKTIDHYDPESEIDIRKVVSCYKLIKECQPEFRSLISELLGVIRCGLRLLK
RSKYQEQARTVSDEDAPLFCLTRSYYQDGYLTPLRAGPRDLINHYIHLRRENPKHFFSP
KHPCYYARLAFNESVCVYRELFDIERLTKMYVEGDYSKEQEKNLQAILSFVKTLDEGKDF
LIEUMYDDLICDPETDIRCT LIEHKDTDLIGRGFTDVFCT

CPn_0166 207594 207962
No robust homolog present in Genebank/EMBL as of 11/7/98
NCLKQYNKSDSIMSESINRSHLEASTPFIKLINLCESRLVKITSLVISLLALVGAGVT
LVVLFVAGILPLLPVLILEIILITVLVLLFCLVLEPYLIEKPSKIKELPKVDELSVVETD

CBm_0167 208309 207977
Nombolog present in Genebank/EMBL as of 11/7/98
NLWSHFPRGFFMLPFCPTILLAKPFLNSENYBLERLAATVDSYFDLGQSQIVFLSKQDQG
INVESELSAKDRKFKPGSMNCTLYTEDPILPAHNSFSNCSDIQMRTPISPIH

CFE_0168 208716 208417
Nd :ZODUST homolog present in Genebank/EMBL as of 11/7/98
SYINLERRENPEHFFNPGHPCYYARLAFNESVRIYRKLFNTAELKQMYGAGDYEQQNEDN
LKSILSFVQILDEKDGFDDFLATHKDTTFIGRGADIFCS

209537 208710

CEMENTORY 208710

NO. ECODUST HOMOCOG PRESENT IN GENEDANK/EMBL AS OF 11/7/98

SEMESTIGENMKNVGSECSOPLVMELNTOPLENLCESRLVKLTSFVIALIALVGGITL

TALAGAGILSFLPWLVLGIVLVVLCALFLLFSYKFCPIKELGVVYNTDSQIHOWFQKQRN

KDBEKATENPELFGENRAEDNNRSARSQVKETLRDCDGNVLKKIYERNLDVLLFMWYPK

TMDDVDPVSEDSIRTVISCYKLIKACKPEFRSLISELLRAMQSGLGLLSRCSRYQERAKT
VSHKNABJFFCPTHSYVBDGVITPLBAGDBVINDBAI VSHKDAPLFCPTHSYYRDGYLTPLRAGPRYIINRAI

VSHKDAPLFCPTHSYYRLGYLTPLRAGPRYIINRAI

LEB
CP1_0170 211098 210025

NOSEDUST homolog present in Genebank/EMBL as of 11/7/98

NVRKNHIIRGEKYNTCTVIAFVLSMSYDTLFKNLEKEDSVHKICNEIFALVPRINTIACT
EAFERNLPKADIHVHLEGT ITPOLAWILGVKNGFLKWSYNSWTNHRLLSFKNPHKQYSNI
FRNFQDICHEKDPDLSVLQYNILNYDFNSFDRVMATVQGHRFPPGGIQNEEDLLLIFNNY
LQGGLDDTIVYTEVQONIRLAHVLYPSLPEKHARMKFYQILYRASQTFSKHGTLRFINC
FNKEFAPQINTOEPAQEAVONLQEVDSTFPGLFVGIQSAGSESAFGACFKRIASGYRNAY
PGGGFGEALAGEFLIFTETIFES SKINDFCILETBUTTESET MENDOECT ALD MOOCH DSCFGCEAHAGEGIETRTIFSSAKVNPEGLIEITRVTFSSLKRKQPSSLPIRVTCQLG

CPN 0171 212444 211149

CPNED171

*GUAA-GMP Synthase
IIKLQSARRHLNTIFILDFGSQYTYVLAKQVRKLFVYCEVLPWNISVQCLKERAPLGIIL
SGGPHSVYENKAPHLDPEIYKLGIPILAICYGMQLMARDFGGTVSFGVGEFGYTPIHLYP
CELFKHIVDCESLDTEIRMSHRDHVTTIPEGFNVIASTSQCSISGIENTKQRLYGLQFH
EVSDSTPTGNKILETFVQEICSAPTLWNPLYIQQDLVSKIQDTVIEVFDEVAQSLDVQWL
AQGTIYSDVIESSRSGHASEVIKSHHNVGGLPKNLKKLVEPLRYLFKDEVRILGEALGL
SSYLLDRHPFPGFGLTIRVIGEILPEYLAILRRADLIFIEELRKAKLYDKISQAFALFLP
IKSVSVKGDCRSYGYTIALRAVESTDFMTGRWAYLPCDVLSSCSSRIINEIPEVSRVVYD
ISDKPPATIEWE

213237 212440

CPn_0172 213237 212440
*impD-Inosine 5'-monophosphase dehydrogenase (C00H-terminal region only)
APIGAAIGIGPLGISRAHHLVEAGANVLVIDTAHAHSKGVFQTVLEIKSQFPQISLVVGN
LVTAEAAVSLAEIGVDAVKVGIGPGSICTTRIVSGVGYPQITAITNVAKALKNSAVTVIA
DGRIRYSGDVKALAAGADCVMLGSLLAGTDEAPGDIVSIDEKLFKRYRGMGSLGAMKQG
SADRYFQTQGQKKLVPGGVEGLVAYKGSVHDVLYQILGGIRSGMGYVGAETLKDLKTKAS
FVRITESGRAESHIHNIYKVQPTLNY

CPn_0173 214041 213715 No robust homolog present in Genebank/EMBL as of 11/7/98 TIFDLIYKLDSYKHQQGFMDFSVFPDRFVESTSPGPIEDIDAKTLVSNCCHYCSRCLFIF LGLLSIIICFSVYGTSGETASLVFGILSLIVLVLLIIECRNRECCRRIS

CPn_0174 214215 214724
No tobust homolog present in Genebunk/EMBC as of 11/7/98
PIFEWMYRKIVILGMIMTTIGNSDSPALMPELSLIPPMTLVGCGTGTGLAYTIPAQGRRS
TERTILDIFTIILGLATIIGTFIVIFFLNGLNCLGTFSILGGGCLITVGLFFLMGLYFM ISSLDQ/SLVGLLQKELSQAEEREEEYIQETEALRGAPRAESFTESPSTWL

214856 215275

CPR_ULY)

LDLACTOPELLRRPDMEOPHCYTQUTTVLYALMSFDPRLSDDTHRLGKOSPLEAENALGE
FIRGLDTHRFFLEEVAIFTLEMYHPKFYLSFIDRDGGVHYEVLDGVFLKTVAACIIENS

FLTDSMSPELLSEVKEALKP

CPn_0176 215271 216518
CT153 hypothetical protein
NDDDPMDESDGEEASKDSAFSAFSYFVKGSTREGKNTVTHGTASRTLYILRQDCSYDP
RALKVDDEPRYWWEKRLDANPDGLNAFVKEVGTHYVASVTYGGIGFOVLKMSYLQVEEL
EKEKISISVAAASSLLKSKTSNATEKGYSSYGSESSAGTVFLGGTVLPDLQDKLDFKDW
SESIENDEFIPLAISVSSITDLIFELFPSEDAQVLSQKKSALGQVTLNYLESHKPKEEGP
STATUS OFFICASIONATION OF A OPENTAL POWNER OF A PROTECTION OF A OPENTAL STATUS OFFICASIONAL OF A OPENTAL STATUS OFFICASIONAL OFFICA

217513 216608

CPn_0177 217513 216608
No robust homolog present in Genebank/EMBL as of 11/7/98
DKREQTKSKFIFLISEESMKQPMSLIFSSVCLGLGLGSLSSCNQKPSWNYHNTSTSEEFF
VHGKKSVSQLPHYPSAFRTTQIFSEENKDPYVVAXTDEESRKIWREIHKNLKKGSYIPI
STYGSLMHPKSAALTLKTYRPHPIWINGYERSFNIDTGKYLKNGSRRRTSHDGPKNRAVL
NLIKSSGRRCNAIGLEMTEEDFYIARRREGYYSLYPVEVCSYPQGNPFVIAYAWIADESA
CSKEVLPVKGYYSLVWESVSSDSLNAFGDSFAEDYLRSTFLANGTSILCVHESYKKVPP

CPn_0178 218052 217789
No robust homolog present in Genebank/EMBL as of 11/7/98
VKEYLDFLVQRNVERDPQTKRHCTVSQKFGGESIDAKTTTGQLFHIAGKTEPGHGKLCLG ESILKOLLALGIITGYENREREVWYYLD

CPn_0179 218550 218056 No robust homolog present in Genebank/EMBL as of 11/7/98 PKIWDTHFETRIEATSVPKFNRRLRKSFHKSGRSSRPSKACVANFFNFTLQAGRSGIIPG KKAILLNYNDAKTPNYSCIFESIGFFNEQDLEAQHNQQAALVRKILKVVPHHFLKGLIAK LPRSLKKDRKFMSSLIFTKLSYALDLSAPMHLEGKPNLSYEEKLD

CPn_0180 218963 218355

No robust homolog present in Genebank/EMBL as of 11/7/98

TSLHKILDCKYRPVFIQNTVASETYPSGILHAQREVRDAYFNQADCHPARANQILEAKKI
CLLDVYHTNHYSVFTFCVDNYPNLRFTFVSSKNNEMNGLSNPLDNVLVEAMVRRTHARNL
LAACKIRNIEVPRVVGLDLRSGILISKLELKQPQFQSLTEDFVNHSTNQEEARVHQKHVL
LISLILLCKQAVLESFQEKKRSS

CPn_0181 219175 218777
No robust homolog present in Genebank/EMBL as of 11/7/98
VHELFKIDGVYYFFKKFMKLFYNNYSLNSHHEKPSSLEKAVQALDSYFYWGGDTTDVLAR
DDISREIYCVRRLYIRFWIVSISQSLSRIPWRLKRILLRYCTLRGKYVMPILIKRIAILL GLIRFSRLRKSVY

CPn_0182 220704 219334
accC-Biotin Carboxylase
RCIMKKVLIANRGEIAVRITRACHDLGLSTVAVYSLADQEALHVLLADEAICIGEPQAAK
SYLKISNILAACEITGADAVHPGYGFLSENANFASICESCGLTFIGPSSESIAMMGDKIA
AKSLAKKIKCPVIPGSEGIIEDESEGLKIAEKIGFPIVIKAVAGGGRGIRIVKEKDEFY
RAFSAARAEAEAGFNNPNVVIEKFIENPRHLEIQVIGDTHGNYVHLGERDCTIQRRQKL IEETPSPILNAEIRVKVGKVAVDLARSAGYFSVGTVEFLLDKDKKFYFMEMNTRIQVEHT ITEEVTGIDLVKEQIHVAMGNKLPWKQKNIEFSGHIIQCRINAEDPTNNFSPSPGRLDYY LPPAGPSIRVDGACYSGYAIPPYYDSMIAKVIAKGKNREEAIAIMKRALKEFHIGGVQST IPFHQFMLDNPKFLESNYDINYIDNLLAQGNSFFKEF

CPn_0183 221207 220695
accB-Blotin Carboxyl Carrier Protein
RRLGMDLKQIEKLMIAMGRNGMKRFAIKREGLELELERDTREGNRQEPVFYDSRLFSGFS
QERPIPTOPKKDTIKETTTENSETSTTTSSGGFISSPLVGTFYGSPAPDSPSFVKPGDIV
SEDTIVCIVEAMKVMNEVKAGMSGRVLEVLITNGDPVQFGSKLFRIAKDAS

221814 221221

efp-Elongation Factor P QWKIKFCCCEEKIMVLSSQLSVGMFISTKDGLYKVTSVSKVAGPKGESFIKVALQAADSD VVIERNFKATQEVKEAQFETRTLEYLYLEDESYLFLDLGNYEKLFIPQEIMKDNFLFLKA GVTVSAMVYDNVVFSVELPHFLELMVSKTDFPGDSLSLSGGVKKALLETGIEVMVPPFVE IGDVIKIDTRTCEYIORV

CPn_0185 222457 221765

TPP/AraD-Ribulose-P Epimerase
nevkKQESYLVGPSIMGADLTCLGVEAKKLEQAGSDFIHIDIMDGHFVPNLTFGPGIIAA
AEVKKQESYLVGPSIMGADLTCLGVEAKKLEQAGSDFIHIDIMDGHFVPNLTFGPGIIAA
INRSTDLFLEVHAMINNPFFFIESFVRSGADRIIVHFEASEDIKELLSYIKKCGVQAGLA
FSPDTSIEFLPSFLPPCDVVVLMSVYPGFTGOSFLPNTIEKIAFARHAIKTLGLKDSCLI EVDGGIDQQSAPLCRDAGADILVTASYLFEADSLAMEDKILLLRGENYGVK

222878 224068

*similarity to Cps IncA
PIKDKILMSSPVNNTPSAPNIPIPAPTTPGIPTTKPRSSFIEKVIIVAKYILFAIAATSG ALGTILGLSGALTPGIGIALLVIFFVSMVLLGLILKDSISGEBERLREEVSRFTSENQR ALGTILGLSGALTPGIGIALLVIFFVSMVLLGLILKDSISGEBERLREEVSRFTSENQR LTVITTTLETEVKDLKAAKDQLTLEIEAFRNENGNLKTTAEDLEEQVSKLSEQLEALERI NQLIQANAGDAQEISSELKKLISGWDSKVVEQINTSIQALKVLLGQEWVQEAQTHVKAMQ EQIQALQAEILGMHNQSTALQKSVENLLVQDQALTRVVGELLESENKLSQACSALRQEIE KLAQHETSLQQRIDAMLAQEQNLAEQVTALEKMKQEAQKAESEFIACVRDRTFGRRETPP PTTPVVEGDESQEEDEGGTPPVSQPSSPVDRATGDGQ

CPn_0187

predicted methylase
predicted methyl

CTH_0188 25090 226406
CTL12 hypotherical plotein
RTLEEX IMPRILEPEXKREGOVORLINNOLLOAT FOTENVLERHBASKEACVLTYYGLL
RTLEEX IMPRILESHEPTSKERIOOVORLINNOLLOAT FOTENVLERHBASKEACVLTYYGLL
RTVFTLVFFERESQHEFTNEMVEWELT KETHOVEKETVA IVEAXYHATESH GEVILVGSF
EVEY WAS LEMESSEESHENTEFTSWEPTELINVENTYFTTELVSSMETET LIVEASHIY ITTO
IMPTOYAKLESGEHSMTALYFTSREVEYLLLYLALFOY YAFLEPMA FORTSALISTLITG
LVWLVFQRAFFSLQVS TENYSETYGALVALPSFELGLYTYTMTYLFCXARITFT IQNROCT

FIFLODKILPSCYLQLITST/ILALTTPQFNESLSPLTAQFIAKQSKVPIGEVSQCLDVLEKEGFLFPYNMSYQPVFNFSELTIKDIADKLLHREIFKKFNPDLGITFIENSFQNIFNQA SKNKENLTLJEIARRIK

TINDER ISMTASGEONQ IOMKLOGH HKSTFY I VEGSSSFIELK PELASALCNOI I PLSTP
TNOGK ISMTASGEONQ IOMKLOGH HKSTFY I VEGSSSFIELK PELASALCNOI I PLSTP
ITSKQ I HATVSYAKI PLDI TKWKH I EITSQAQLPEVA I HPKDPNLALQLRDTKLGI KKTE
KFSDIRYSSSTVLOGASPSHLNGLI SI DNKKHLTKFRLQQAQLPHTYLRAI FPQPFV INV
PLDVAYYSLNI EGTYKNAHLEADA I LDNPLLKLSCSMSGTWKNFLFKGGOTYHFNKKWQE
ILSPHFSYAEARFSGKAQ I TDTNLFFPKFSGKITARENELLI HAKFGSPNEP I KPETIS I
LIHQOFCSLPLSLVSNHLAPFHLKKLTFSFHTDGKFVTKGNLQALI ENPDY PDLANTRI
LIPDLLLSLDESSTSPSSKDLKI QOSGEIFSL PLDS I TKTYGKQVRLSPYFGSSGDLNFV
VNYNPKDQNKLTLLSNFKSEALLGELKLWDFSMKLSSCTOGTLOWEVSPERYASFFKNA
SCSPTCLLHRTANVRLDI SKLSCPEETKGLSCLTLLAAGGLEGSLEATPLIFYDNVSKET
FI INDFKGSLRANNLDAKI EYDLKGSCLAPRODSKTLAEFSLEGQVDHLFSPESREFKQT
TANWIHIPSSFI AGI I PMSPGLKAQ I SSLAGPRINVSIKNAFRFGEGPVDI IMVDSENLQAQ
IPLILNEKSILLRENLTAHLSINEDVNKAFLOEFNPLLAGGAYSQY PVTLEIDKQNFYLP
IRPYSFEEFRIQSATLDMGKIS I ANTGTMYALFOFLDITDQKQFVESWFTPI FFSVQKGS
ICKRLDALI DRRI RILLWSKTD LAHDRLFMTLGI DPEVIKKYFHNTSLKTKNFFLIKIR
GSISSPEVDWSSAYARIALLKSYSLGNPFSSLADKLFSSLGDSTPPPTVHPFPWEKSNFD GSISSPEVDWSSAYARIALLKSYSLGNPFSSLADKLFSSLGDSTPPPTVHPFPWEKSNFD

CPn_0190 229901 231274
No robust homolog present in Genebank/EMBL as of 11/7/98
LLGIKLMRKRHSFDSTSTKKEAVSKAIQKIIKIMETTDPSLNVETPNAEIESILQEIKEI
KQKLSKQAEDLGILEKYCSQETLSNLENTNASLKLSIGSVIEELASLKQLVEESIEESIG
QQQLIQSVLIEISDKFLSSIGETLSGNLDMNQNVIQGLLIKENPEKSEAASVGYVQTLL EPLSKRIGETHKKVATHDVNISSLOFHMMSVAGGRFRGHIDMNGYRVLGLGEPKNGEDAV SKDYLERYVSSQLTIDKVEDKPITKPNKGKLLYSQGTSPKLESPLPLGLLTSGISGFTWK SASKSNDGSFPFSALRHKETESDTDCFQITSTTLSGNQAGTYTWSLSLKVLVPSIFQIEK PEVQLSLVYSYEDWLPIDNIFNMSQPRTIPLALLGQTMLAGQKYDILELAAHQTNQTLMI SPNCSRFSLQLKQTNQFENSPVDFYIVHAAHSCHWSGF

CPn 0191 232039 231314 ging-ABC Amino Acid Transporter ATPase OHRFPVFLGYOKREGYMTIRVRNLAYSVNKKKILDGVTFSLERGHITLFVGKSGSGKTMI LRALAGLVOPTOODIWIEGEAPALVFOOPELFSHMTVLGNCTHPOIHIKGRSTEEAREKA FELLHLLDIEEVAKNYPDOLSGGOKORVAIVRSLCMDKHTLLFDEPTSALDPFATASFRH LLETLRDOELTVGLTTHDMOFVHSCLDRIYLIDOGTVAGVYDKRDGELDSGHPLSKYIHS YO.

CP010192 232643 231984
ghn-ABC Amino Acid Transporter Permease
EvgyDHWLAIARLLLRCGGYTLCVSGIGILCGSILGGLIGTVTSLYFPSKLTKLLANSYV
TVIRGTPLFIQILIIYFGLPEVLPIEPTPLVAGIIALSMNSAAYLAENIRGGINSLSIGQ
WESAWLLGYKKYQIFVYIIYFQVFKNILPSLTNEFVSLIKESSILMVVGVPELTKVTKDI
VSRELNPMEMYLICAGLYFLMTTSFSCISRLSEKRRSYDN

CPn_0193 233144 232686

##IGHT9 233144 232686

##IGHT9HEALTH 232686

Cen_0194 233162 234241
gdp-O-Sialoglycoprotein Endopeptidase
EVPHTIKGNVFFSWFFMLTLGLESSCDETACAIVNEDKQILANIIASQDIHASYGGVVPE

LASTAHLHIFPQVINKALQQANLLIEDMDLIAVTOTPGILGSLYGVYPGKGIAGAYGGVVPG LASTAHLHIFPQVINKALQQANLLIEDMDLIAVTOTPGILGSLYGVYFGKGIAIGAKS LEGWHVEAHLYAAYMAQNVQFPALGLVVSGAHTAAFFIENPTSYKLIGKTRDDAIGET FDWGRFLGLPYPAGPLIEKLALEGSEDSYPFSHAVVPNYPFSFSGLKTAVLYAIKGNNS SPREPAPETSILEKQRDIAASFQKAACTTIAQKLPTIIKFSCRSILIGGGVAINEYFRSA LQTACNLPVYFPPAKLCSDNAAMIAGLGGENFQKNSSIPEIRICARYQWESVSPFSLASP

234172 235785

CPn_0195 234172 235785

oppa-Oligopeptide Binding Protein
YSGNSYMRKISVGICITILLSLSVVLQGCKESSHSSTSRGELAINIRDEPRSLDPRQVRL
LSEISLVKHIYEGLVQENNLSGNIEPALAEDYSLSSDGLTYTFKLKSAFWSNGDPLTAED
FIESWKQVATQEVSGIYAFALNPIKNVRKIQEGHLSIDHFGVHSPNESTLVVTLESFTSH
FIKLLALPVFFPVHKSQRTLQSKSLPIASGAFYPKNIKQKGWIKLSKNPHYYNQSQVETK
TITHFIPDANTAAKLFNQGKLNWQGPPWGERIPQETLSNLQSKGHLHSFDVAGTSWLTF
NINKFPLNNMKLREALASALDKEALVSTIFLGRAKTADHLLPTNIHSYPEHQKQEMAQRQ
AYAKKLFKEALEELQITAKDLEHLNLIFPVSSSASSLLVGLIREGWKESLGFAIPTUGKE
FALLQADLSSCNFSLATGGWFADFADPMAFLTIFAYPSGVPPYAINHKDFLEILQNIEQE
QDHQKRSELVSQASLYLETFHIIEPIYHDAFQFAMNKKLSNLGVSPTGVVDFRYAKEN

235906 237519 CPR_0196 235906 237519
oppA-01igopeptide Binding Protein
KLKSYSKERSFMLRFFAVFISTLWLITSGCSPSQSSKGIFVVNMKEMPRSLDPGKTRLIA
DQTLMRHLYEGLVEEHSQNGEIKPALAESYTISEDGTRYTFKIKNILWSNGDPLTAQDFV
SSWKEILKEDASSVYLYAFLPIKNARAIFDDTESPENLGVRALDKRHLEIQLETPCAHFL
HPLTLPIFFPVHETLRNYSTSFEEMPITCGAFRPVSLEKGLRLHLEKNPMYHNKSRVKLH RETUGETER OF LEARNING TO THE ARMY STREAM TO THE STREAM HKKHLRNQLIIEALDYLEHCHILEPLCHPNLRIALMKNIKNFNLFVRRTSDFRFIEKL

CPI_D197 237512 19882

OPA_OTIODEPTI O BIDDITE PLOTE II

MYRHKELIDLEFIDIEFI EVI FEKMERRETELELELITE ODYOCKHKODLI I PIHDDPV

AFOFFOAKRAMULD LAOLLEFOLTRETHRESHDELLA I ACRYTVOEDFOSYTTFI KDSAL

WICOTTE TODED I KNAMEYAQENDENI O FROGUNETETE SANAIT HILDSENPEPEKLLAFP

AFA I FEKENPKLESOLYTEVEY FESHI TILLKRENTYYDYHCVSINSI KLLI I PDI YTA IH

LUNGCKVIMVOOTIMOOTI PMELIKOSON HYYTT PVEGAPMICLINTKSPHLINDLONRIRLA

TOTUKKU I HEALGOTOOPAETILEROATOOPAGYKKOKPUTEOEKUULTYPEDI LIRCOR IA

ETILKEOMKAAGI DULLIDELEYHLEVNKRKVODYATATOTOVAYYFGANI. IDEEDKLLONF 237512 138882

EIIPIYYLGYDYLTQDFIETVIYNAJGAVDLKYTYF:

2:9145 CPn 0198

241018 241983

CPH_0199
appB-01igopeptide Permease
kCLIGLSLVFSYIKNRILFNLLSLWIVLTLTFLVMKTIPGDPFNDEGCNVLSEEVLQTLK
SRYGLDKPLYQQYTQYLHSIAKLDFGNSLVYKDRKVTNIISTAFPISAILGLQSLFLSIG
GGIALGTIAALKKKKQRRYILGASILQISIPAFIFATLLQYVFAVKIPLLPIACWGSFTH
TILPTLALAVTPMAFIIQLTYSSVSAALNKDYVLLAYAKGLSPLKVVIKHILPYAIFFTI SYSAFLITTVITGTFAIENIFCIPGLGKWFICSIKQRDYPVALGLSVFYGTLFMLSSLLS DLIQSIIDPQIRYAHGKEKKRK

241996

oppC-Oligopeptide Permease EKKKHKLMENLSSAPSRSIWKSIIQNKMLVLGLTTLIILMLGALLLPWFYQDYEQTSLKD EKKNAKLMENISSAFSKS I MAS I (INNILVISUEL ITTELLEALELFFF (1915)AUSTUMPS I LIVSPCSRFPFGTDTLGRCMFARTLRGLRLSLLIATIATLIDVCVGLLMATVAISGGKKI DFLMMRTTEILFSLFRIPIIILLLVIFHHGLLPLILAMITGWIPISRI IYGGFLLLKNK PFVLSAKAMHASTFHILKKHLLPNTLAPIISTLIFTIPNAIYTEAFISFLGLGIQPPQAS LGTLVKEGINAIDYYPWLFFFPSLIMIALSISFNLIGEGAKTLCLEEGSHG

CPn_0201 242810 243715
oppD-01igopeptide Transport ATPase
ASISSARGIKHYVSKRDLMDNYLLNIKDLTITSTNPKRTLIENLSLQLKENRNLALVGES
GSGKTTITKAILGFLPENCLIKTGSILFEDIDITKLSPKELHKIRGQKIATILQNAMGSL TPSMRIGMQIIETLROHHKMNKEEAYNKAMQLLTDVCIPNPKYSFSQYPFELSGGMRQRV VIAIALASQPKLILADEPTTALDSMSQAQVLRILRNIQQQKQATILLVTHNLSLVKELCN DICIIKDGKLIETGTVEEIFLSPKHPYTLKLLNAVSKIPIKKTSSPILKNKFQPLMSMQG

CPn_0202 243682 244500

oppF-Oligopeptide Transport ATPase

vPTSNEYARWFMTTLLSIKDLSLTIRGKKILNHINLNLIKGSYLTIVGPSGSGKSSLALT

LDLLKPTTGTIFHMDPKIFRARKVQVIWQDIDSSLNPCMSIKGIISEPLNIIGTYSKA

EQNKEIYNVLDLVNLPKSVLHLKPYKLSGGGKQRIAIAKALVSKPELLICDEPLSSLDTL

NQSLILDLFQTIKKEYQNTLLFITHDMSAAYYIADTIAVMDQGSLVEHACREKIFSTPKH TTTQDLLDAIPIFSLISTEMEPSEEYELQVASK

CPn_0203 244966 245802

No robust homolog present in Genebank/EMBL as of 11/7/98

IVPLEOKNNKETSCHMTYTFSPTLOKSFSLFLLEKLDSYFFFGGTRTQILVITPTNIRLA

AKKRGCKVSTIEKIIKILSFILLPLVIIAFILRYFLHKKFDKQFLCIPKVISNEDEALLG

SRPOAVEKAVREISPAFFSIPRKYQLIRIDTPKDDAPSILFPIGIEIILKDLCIDTLKQS

NLFLKREMDFLGHPEEKALFDSICSIEKDQEWMSLESKKLLITHFLKYLFVSGIEQLNPG
FNPENGRGYFSEISTAKIHFHQHGRYGPIRSSGPIMKEI

CPn_0204 245691 246002 No robust homolog present in Genebank/EMBL as of 11/7/98 PREWAWVFFRNKYSKDPFSSARSIWANPFFGTHHEGNIKIKGMGYQIFTRLKKLGISFSS YNSINPNPYFFDEGCFVYWESQFKSALQDHGILQKQTETFYRNT

246073 246327 CPn 0205

No robust homolog present in Genebank/EMBL as of 11/7/98
IEDSIKGYGSASAFRNPPQLLLKFFLVCEELCILTVATHRALLETPLALSFFKELKTKYV YRAKDILQLHNYKGFTILNTSPLCS

0206 246346 247161

CMT_U2Us 24/161
CT203 hypothetical protein
IVDRRSPACYDSINSDAIGYSLLMDISHILEDLAYDEGILPREAIEAAIVKQMQITPYLL
HILHDATQRYPEIVNDGSYQGHLYAMYLLAQFRESRALPLIIKLFAFEDDTPHAIAGDVL
TEDLPRILASVCNDDSLIKELIETPKINPYVKAAAISGLVTLVGAGKIPRDKVIRYPAEL LNYRLEKQPSFAWDNLIAGICTLYPGELFYPISKAFDGGLVDTSFISMEDVENIIHEETV ESCIHTLCSSTELINDTLEEMEKWLEDFPIEP

CPn_0207 247208 248617
ybh1/sodit1-Oxoglutarate/Malate Translocator
vnkkkrflsllfltavlljiwfsphpasinsnawQlfaiffttimGiffQPvpmgaIAII
GISTLLLTOTLTLEQGLSJFHNPIAWLVFLSFSIAKGIIKTGLGERIAYFFVSALGKSPL
GLS'GLVITDFFLAPAIFSVTARAGGILYPVVTSLSDSFGSSAEKGTODLIGSFLIKVAY
QSSVITSAMFLTAMAGNPLVAALAGHVGVSLSWVLWAKAAIIPGLPSLFLMPIILYKLYP
PKITOCEBAIRSAKLRLKEMGPLKKEEKTILMIFFLLVVLWTFGDLLGISATTAALIGLS
LLILTNILDWQKDVIANTTAWETFIWFGALIMMASFLNQLGFIPLVGDSAAALVSGLSWK
IGFPLLFELYFYSHYLFASNTAHIGAMYPIFLAVSISLGTNPIFAALTLAFASNLFGGLT
HVGSGPADLYEGSHLUTNINDEWM SGEPAISINUTUMTGIGLIMWKALGLI HYGSGPAPLYFGSHLVTVQEWWRSGFALSIVNIVIWIGIGSLWWKALGLI

CPri 0208 248935 250602 CPT_0208 2-4935 250602
pfkA-Fructose-6-P Phosphotransferase
SVAVILMBPLYVDLDTIISSYSPPLPKEROEAASLIAVPDTSHSKPVVPGVKTLFPQTYH
LPYLKFVQGENVVHTDLK\GVMFSXGPAPXGHNVIGGLFNSLKDFHPDSSLVGFVNNGDG
LTTBIKSIDITEEFLSKFRN.%GFNCIGTGRKKIVTPEAKEACLKTAEALDLDGLVIGGD DITIONS TO TEST SERVING FROM THIS RECOVERABLE DELYARADED DO TO THE GENERAL ALLEST PARKENTS TO YOU PYTHOU DO DO THE FORD THE STATE YES TISSUES DALOCKARY HEIKEMERS AS HEALE CALOTHEMENT DESCRIPTION THE SERVING THE PWCPHOCIETEDPTTY

CPH_0209 251909 251272
No robust homolog die ent in Genebank/EMBL au of 1177/98
NDJUNIRLKMETYSETTELSKITCL/IMERILD/YFCFY/KETPVIA PTPAGLAIA/EQNI

CPn_0210 252459 251440
No robust homolog present in Genebank/EMBL as of 11/7/98
YQKLWEREREYFKTIREKEHATISTMLVELEALKREFAHLKDQKPTSDQEITSLYQCLDH YQKLWEREREY KTIKEKEHATISIMDELEALAKE AHARDYLAGI SOOFTI TETUG LEFYLLGLOODKFIKATEDEDYLFESQKAIDAMIALLIKARDYLAGIGIIGAIYQTIEFLG AYLGKYNRRAFCIAGEIHFLKTAIRDLNAYYLLDFRWELCKIEEFYUWSNDOVEIAKRKL CTERKETKELMESLLREEHAMEKCSIODLORKLODI TIELHDVSLFGFSKTPSQEEYQKD TWALIBERGIIJAH TIAGIT TÜRLÜL SOOTTIRITÜ

CPn_0211 252765 252463 No robust homolog present in Genebank/EMBL as of 11/7/98 ECVMSYPDISNYQASSIQSALLHKTSDQIQQKRCFKQSTFVILAVSLVIIGSLFLLAGVA ILTVFSHGVLSLVFGVLGIVLGLLLLAGGVGLLVEEAKSLL

CPn_0212 254081 252888
No robust homolog present in Genebank/EMBL as of 11/7/98
ELSYGVIVSIYSEILSFSELTSCKHSLFPFGPIETASIRIHHVFNIVVIVCLIILGTLFVC
LGMVFLGVFSTYLLGMSSMILGLLLISIGLALLKFKERYGLEPKELFGVEGGFDKKLPSE I IQMQQIADLARELDLEQKKDTLIRGFSARLDVLEGSKTEKKQILKIGVPRNLSETQER
AQEONS ILEQCKEALLFRRKSAQEIFKKLYDRKAAFWRSYREDLWCYSEIHVSKKALSNL
YIGDVFEGTAPHFLMEAYAMCRTAKNLRNYVKVCVEDMRVNEEKKRAKQLSVSELLCCCT EIETDLENETNLFTSDSEDVLEEYQIHCIRVTMLHALWAIYNDEVVSRKPIDTLDRVRAR MAVEDCIETFEELQMCVVHTKTLELEIAQLYVDILLEA

CPn_0213 254345 254190
No robust homolog present in Genebank/EMBL as of 11/7/98
ILVVFSRVIFSNTNQIGIPRLELILPLWKKENDPFCFLFSRVEGTFIILNIK

CPn_0214 255768 254446
No robust homolog present in Genebank/EMBL as of 11/7/98
FLGLKEDYERPTYCNIPPAPHPQRVDSKGCIASHVSTVVVVALFILGIFFLSGSLAFLVH
TSGGVLLGAALPILCIGLVLLAVALIVFLCHKHKTRQDLDYYDQDLDSLVIHKKEIPNDI TSCGVILGAGLITIC IGLVLLAVALIVE ICHREKTRODLDY I DEDISOVIRE IMBREI PROLI SELRVITERELQNLFGOFTKOPSOLSQELQGKF, INCMERWILTLEDEVTKFLIVRORFLETR RNFTTFGEQVKGIQSNIFDLHEEKSSLYLELYRLRKDLQVLLNFFLLPGILKVDYDEIE AIKGLFIRLTSALDKLDVKAQERKKFINEMSREFKEVEKAFDIVDRATKKLMDRAKKESP ARLFMGRTESLLEMKKNEEALKNQGLDPENLSHPELFSPYQQLLILNYLNSEIVLHYYEF LISGTVTSGLTLEECENRMRAASTGLNALLVRKLQFRGAIKSAYFEKLTEIEKELRSLQD VIKSLELELIHKIKDIVTEET

CPI_0215 257039 255759
No.ipobust homolog present in Genebank/EMBL as of 11/7/98
LTSSKKQVMSSAIARDCFPSPSPQPSSTLGVHPPKYKSLILSVSLIVLGVLLLCVGMLLL
VM##FSFSVLTVGLGSAGVFLGSLLLILGLIFFVSYHRKLSEATRSLEGKITLEYQFWAD VNRFFSTSVITVGLGGGGVFGSSLLLITGLIFFVSYHKKLSEATRSLEQKITLEYQFWAD LRKELNEVQEWSNFLLDEWEDFKEVVAQHKSQFATFEGDLLLFGREVEKYETIKKELDOR DWAELTELKNIWGPLEFLRKKGDRLQCEIDKLRKEVMKVGKSGLKLACELTKFKSALKDV KIEQECYRDKRKVEKLEVFPEGYRRELLEVLKTRLSVEKEIQLFEEVVSAFEEKLASLHR TVFSEEELQEALDKAKAELLDIQVRKSVVEDLSCEPTLIQYHLLRLYEVQCRIVEQFLTQ TFSSEQEKVLEEYEALKARIRKTLRVKLDQVRANVAFVASTTDLLSESESLDGNDSVFED

CPT D216 257623 257174

No pobust homolog present in Genebank/EMBL as of 11/7/98

NKARTMNPVTFDRIQVDFIPEDTSLRINSYIVAGGLLILGVVLSILSVICLDIGLVGLSA

GAAFTLGLGCLIFALFIFFSFSLILLLSQEKRVPDVLSLYLEKEVPQYETPLYKEDLESER DMSAISERLGIIEEKLRIAEKFRYSDSVFV

CPR_0217 257881 258579
ypäEP
PKCKLKGFLSVNELIFGFQTFSVVVLGVFFASRGKAWLTGWLSLLSSIMNVFVLKQIHL
WGFEUTSADVYVIGLUTCLNYAREHYEKNDINDAMLCSWVISIAFLVLTQLHLFLIFSPN
DS\$QEHFLALFSSTPRIVVASLVTLIFVQIVDIKLFTFLQRVFSKKYFAMRSTISLLFSQ
LITHTIFEEDICIVGUUDG DDBJATAN JVGUUDG TENDER TENDER LIDTIIFSFLGLYGLVSNLCDVMIFAMLVKGIVITLAIPTLTVTKAVLDRRSS

CPN-2218 259064 258582

Nolfgbust homolog present in Genebank/EMBL as of 11/7/98

ILLSKKVFFESYEDFANVASSWPKSLRALVQGRYFVDSELKETPYRIHDFKKTPIHHRLY
RSLPIISTIGGIIRLIEAHSGPIHPRDKMKYRFEVLQAVIEILGLGVLILVFDIIGCFLA
FLVAIILSLLLYCNSTFTCVQNLSFTERMLEGIGEAVNFLA

CPn_0219 259348 260472
tgt-Queuine tRNA Ribosyl Transferase
GSSLALKFHLINGSKKSQARVGQIETSHGVIDTPAFVPVATHGALKGVIDHSDIPLLFCN
TYHLLLHPGPEAVAKLGGLHOFMGRQAPIITDSGGFGIFSLAVGSVAEELKSGKKKGMS
SLVXITDEGAWFKSYRDGRKLFLSFELSVQAQKDLGADIIIPLDELLPPHTDQEYFLTSC
SRTYVWEKRSLEYHRKDPRHOSMYGVIHGGLDPEQPPIGVRFVEDEPFDGSAIGGSLGRN
LQEMSEVVKITTSFLSKERPVHLLGIGDLPSIYAMVGFGIDSFDSSYPTKAABHGLILSK
ACPIKICONYSODSEIDBEGGGITGLEGLEDVERIK AGPIKIGQQKYSQDSSTIDPSCSCLTCLSGISRAYLPHLFKVREPNAAIWASIHNLHHMQ QVMKEIREAILKDEI

CPn_0220 260660 261238
No robust homolog present in Genebank/EMBL as of 11/7/98
FYSFLKKKGIFYMSKESIRSYSEISTPTPFIFRETPSYEGVAYKLQLRSPAKDCILRNRVS
LKGALLRSIPPYGSFLGAKRIHSAWSAKDAPCTTRV/HYLVGGLELLGLGVVVLACKVLA
TAŁKFLFSKASSKIKQMKWREKARNLAAKDTVQSIKEFCSVDLTSCFTRCFRLRNRVVEE

CPn_0221 261621 262064
NO tobust homolog present in Genebung/EMBL as of 11/7/98
DALRYKYETGIOMYNRYKGSAEFSADHYYDDDIAYRM/KRNLRGLAPVENEVCLFEENNL
LESYMAGIFIMGGILGLGREHSVWSTODFKDGKEGIIFHTALGILETEAGIIVLLIKIT TTILLILFTFCLLCYFMYSAAYSDFHPI

262474 262842

CTM_U2ZZ 502474 20249.
**Weak Similarity to Bacherlophage THPI (Ott4)
GEKFLKIWEKLRELNAFELTOPEEYRNRW/LMTGL/CPFC/RTQHAKVW/YRCVHEASLYE
ENGFUTLTYDDKHLPQYGGLVKLHLQLILKFLRKMIGPHKIRYFE/GAYGTKLQRPHYHL

CPn_0223 262950 263333 No robust homolog present in denebook EMBL as of 11/7/98 "TMLIGRYGGDDGTEATKNTFTIIKLGFVRONLEGLTNPISEIVSETSSSIKDSVLRSL PILGGILGGARLYGTLGTNDPLDETGEKIWHTIFGALETLGLGILILLFKIIFVILHGIF HLVIGFCK

CPn_0224 263402 263674
No robust homolog present in Genebank/EMBL as of 11/7/98
YTFKNPKKNKKKKPNSIIFLENTKHYPDIFREGFVRDRHGLMEASDWLLSTEITIIRSIL Highly Sobrasiwes SECOND SECTION OF

CPn_0227 263846 264541
No robust homolog present in Genebank/EMBL as of 11/7/98
NSFTIKFLLMTKNAINSOTTTPOPNLTDAEPIASRAÇCKSIAVIISLFALGMLLLCLGII
LISIPIPGLAAQVALGLGIVSLIGIALANIGFLCLLLRCKQYPOKPDTLPSESSKOPSE
GSTFTALPWQAGEFLEKVQVSATPILLPKNKDEELSAKVMKEGAEAASSIKQAVLESTEK
LIDARKQEESRREARKKIVAEEAEASRKRIQQQMAADQEALRKRKEEVAKRK

CPn_0226 264545 264967
No robust homolog present in Genebank/EMBL as of 11/7/98
AIFNRKRMPYYANTLEFIQOTQSLCPLFKYGFVRHHYKGQLEIEDASHDWDFLEPPSTWK
RTLLAAIPILGSVIGLGRLFSIMSIREPQDSQEYKSIFWHTLCAVLEILGLGIVALILKI LATFIMAMPGLKRVATFLFYS

CPn_0227 265467 265009

CMI_UZZ/ Z55009
dSbB-Disulfide bond Oxidoreductase
KERRNIFYSCKLLKEIMMINFIRSYALYFAWAISCAGTLISIFYSYILNVEPCILCYYOR
ICLFPLTVILGISAYREDSSIKLYILPOAVLGIGISIYQVFLQEIPGMQLDICGRVSCST
KIFLFSYVTIPMASVVAFGAIVCLLVLTKKYRG

CPn 0228 266242 265412

dsbG-Disulfide Bond Chaperone VKDRADFLNLKEKFSCSILKKENAFEFYVFCSIKOLTNSSLRGPLNKKILVLCTAMFFIV VADIADI BILLEN SUSTINAENT FOR TVYGGI MUDI INSSERENAELT EN FET EN

CPn_0229 266163 267560
CT178 hypothetical protein
NSKAFSFLRIEGENFSFKFKKSALSFTYNTANLTKSTFTFILLLLLRKKDQGLRFMDKET
LENIYRHFRYRFLKLNILPAFLGLLLLCSPNTLNYTQVDVIFSDRLCSCLLIFLAIASLT
KRSLLWLGAPLGIWVTLFACVAGRSPTIFANDTLIGFAILAVVCISPTRPEALEVGFTLP
EGFSYNPSAGGRRAAVLFLSLLGWLEARYLTASSLGITSSQSSNFLLLYSSIMTVYSLLV VLSLAGSERRMHTRPKIVIATALALITGVIILTLLPIILHQLRYDCWLCLCLTIEPALAVV FAYDETRATLRYISQFLGDKRALTRASFFGSEYYKHTLSWEERTVLPLRKAYKQAFEGIS FPINOLLAILVATVFVKVNSSMGLPTFPRNFLNICCWFIIVLFILAFAESLRHLRWMNLI FSAAILFSPVLFHIPVESPMFLPIIVTGLILIILSIGKRRRTKRKL

CPn_0230 268277 267576
CT179 hypothetical protein
RFKKALIYMSSQPLVTTSSSLSRYVVLTGEEKVACYKKAFNHIWHGAPAIILAAALLMFC
IFGFVLGSILLGAPLEGASILYDVILPMLLPSILVFVLLVLPLNIYAYSHHKQVLALHER
ITOSNYKEIYDHCEKEKKPPNKKALSLYTESQVLVPEYSKRFSSNILGKTLKIIPKKDSP ESLKHDELIQKALERAKENIYMNKNQREKRDEREAKKEAKNASKTNPLWEGLGT

CPn_0231 268996 268253
tauB-ABC Transport ATPase (Nitrate/Fe)
PQAFVSIQDRGFSMLQAHRLCYSCDNQVILKDASPQASPGTITIILGSSGVGKTTLFRLL
AGFLPLQEGELLWNGSPLNRKDVAYMQQKEALLPWRTALKNMTLSTELGINTSHNALSNE
RLEEIIHNFDLGQLLDRYPDELSGGQRGRIALAAQCLSLKPILLLDEFFSSLDVLLKEQL
YQDIVALAKKENKTVLLVTHDFHDVSCLGDVLYVIKNKTLTPVPLDPSMRPLNNGLCFIK DLKKHLYT

CPn_0232 Pn_0232 270134 269232 similarity to 5'-Methylthioadenosine/S-Adenosylhomocysteine

Nucleosidase

KKFLMRRFLFLILSSLPLVAFSADNFTILEEKQSPLSRVSIIFALPGVTPVSFDGNCPIP

WFSHSKKTLEQGRIYYSGDSFGKYFVVSALWPNKVSSAVVACNMILKHRVDLILIIGSCY

SRSQDSRFGSVLVSKGYINYDADVRPFFERFEIPDIKKSVFATSEVHREAILRGGEEFIS

THKQEIEELLKTHGYLKSTTKTEHTLMEGLVATGESFAMSRNYFLSLQKLYPEIHGFDSV SGAVSQVCYEYSIPCLGVNILLPHPLESRSNEDWKHLQSEASKIYMDTLLKSVLKELCSS

CPn_0233 270439 270248
No robust homolog present in Genebank/EMBL as of 11/7/98
EKARTMFLGKVLLFLLRISRRSYVQEIGIFFHLETPDLKIVLCAFVSTFIVVEMDVSLKN

CPn_0234 271246 270548
CT181 hypothetical protein
FIMLOSCKKALLSIVVSILAFHPIPGMGVEAKSGFLGKVKGWFSKKEIQEEARILPVKDS
LSWKRYDYTSSSGFSVEFPCEPDHSGQIVEVPQSEITIRYDTYVTETHPDNTVYVVSVWE
YPEKVDISRPELNLQEGFSGMMQALPEDQVLFMQARQIQGHKALEFWIVCEDVYFRGMLI
SVNHTLYQVFMVYKNKNPQALDKEYEAFSQSFKITKIREPRTIPSSVKKKVSL

CPn_0235 271395 272177
kdsB-deoxyoctulonosic Acid Cynthetase
VFVRYLLMKPEESECLCIGVLPARWNCGRYPCKPLAKIHGKSLIQRTYENASQSSLLDKI
VXATDDOHI IDHYTDFOGYAVMTSPTCCNCTERTGEVARKYFPKAEIIVNIQGDEPCLNS
EVVDALVQKLRSSPEAELVTPVALTTDFEEILTEKKVKCVFDSECRALYFSRSPIPFILK
KATTPYLHIGYYAFKREALFRYLOHCCTPLSDAEDLEQLRFLEHCGKIHVCIVDAKSPSV
PVDFLAAVPOV KDWLENAME DYPEDIAKVEQYITCLENAYE

272138 273766

VROWEGK IAVAKFOREQGIPYFGICLGMOVLVVEYARNVLNLDQANSLEMDPNTPHPIVY VMEGQDPLVATOGTMRLGAYPCLLKPGSKAHKAYNESSLIQERHRHRYEVNPDYIQSLED HGLRIVGTCPPQGLCEILEVGDHPWMIGVQFHPEFVSKLISPHPLFIAFIEAALVYSKDA

CPN_0237 273741 274214

yggF Family

SILRMOAMSKPSSCKAYLGIDYGKKRIGLAYAAEPLLLTLPIGNIEAGKNLKLSAEALHK

LICERBIT MITTERIE DE MAR ANDERS TO A FOR A SOUTH OF A MOOPE INVOAERM

LECTOR GURRO PROPERTY A PAYLOR TO A DEPKINT

274210 275838

CPn_0238 274210 275838

zwf-Glucose-6-P Dehyrogenase
PCNHOKLRDFNFRNFLLFVIFASACTKKEIKMTNVVQETIGGLNSPRTCPPCILVIFGAT
GDLTARKLLPALYHLTKEGRLSDQFVCVGFARREKSNELFRQEMKQAVIQFSPSELDIKV
WEDFQQRLFYHRSEFDNNMGYTSLKDSLEDLDKTYGTRGNRLFYLSTPPQYFSRIIENLN
KHKLFYKNQDQCKPWSRVIIEKPFGRDLDSAKQLQQCINENLNENSVYHIDHYLGKETVQ
NILTTRFANTIFESCWNSQYIDHVQISLSETIGIGSRGNFFEKSGMLRDMVQNHMMQLLC
LLTMEPPTTFDADEIRKEK IKILQRISPFSEGSSIVRQQYGPGTVQGVSVLGYREEENVD
KDSRVETYVALKTVINNPRWLGVPFYLRAGKRLAKKSTDISIIFKKSPYNLFAAEESRC
PIENDLLIIRIQPDEGVALKFNCKVPGTNNIVRPVKMDFRYDSYFQTTTPEAYERLLCDC
IIGDRTLFTGGDEVMASWKLFTPVLEEWDQDSSPSFPNYPAGSSGPKEADALIERDGRSW
RPL

275863 276672

CPH_UZ39
devB-Glucose-6-P Dehyrogenase (DevB family)
KSISMTNIGIETMATLINFNDTNKLLLTKOPSLFIDLASKDWIASANQAIKQRGAFYVAL
SGGKTPLEIYKDIVINKDKLIDPSKIFLFWGDERLAPITSSESNYGQAMSILRDLNIPDE
QIFRMETENPDGAKKYQELIENKIPDASFDMIKJGLEDGHTLSLFSNYSALEERDNLVV
FNSVPHLETERMILTFPCVHKGKHVVVYVQGENKKPILKSVFFSEGREEKLYPIERVGRD RSPLFWIISPESYDIADFDNISSIYKMDIL

CPn_0240 277861 276698

No robust homolog present in Genebank/EMBL as of 11/7/98

LVYFMVFSPSSESVVKANSVVRSNFCYFLENKFVSPSESTEVMFSEIMKGRVPDIESLFD

RPTDMMTGFKAAQNLGNLFNSFGILIMCFSQCKSCQTPEKETSAIVLGATLLFFVVALI

LGPTLGALVYCAYKVYTLGKNIYSLNKAKAKVLRHPAQNVFHRAAGVATIRSSEEAVKAC

KLYKSAMIGSLVVSLIASLALIALTAGTVLVLFFVAPGAAPVITVAAMMGCCAAGGGALLI

SLLGLWIAIVRKAKDGEACVGHLTNVVLHTAVSEALLHDPSHFQTNALARDLFLTDCLSH

YGHLFSNEEVAQLVQGGAPGGGSRPSQHYGGSSDYQNRRGGNGNFGGSHFGGGGGFAGSH

FGACYPTAPTMPSAPPPFPPPAYDTIYG

CPE-0241 279372 278203

No probust homolog present in Genebank/EMBL as of 11/7/98

IFLYKFMSAMISLSSSHEASIASNTQVROVLVSLAMDEFVEHNTEILPIKVFLARGTLSS
TATIDDLKDVVETEGEHHFQVYSNISLKHIYQRFFEKIFGIGCCPLLLVTDSHHTDPGA
LITGIFAAVLFTVLAIVFGPTLGILCYSAYKIYQLTKKISSLSRTHTEVINSVQKSDPFI
HRSGAVAAAAASQSTIKACKVFRQSTLIFFVLGLITIISLAALIVGLVFALFFLDPGAPA
WMTÄMIGCCAAGGTGILLSVIGFLLASVYSVQKSQBGVHHMHTALLGIVSNTIIQMPY
LPTGPTKKVLTQSIRRYQOFFSDDEYRDIESEVPLNRQTTPPPSYETLFHEEGSDGSSN
WIDDRSPDAYSTINSGSPDPVVS VIPRESPPAYSTIDSSNSPFPSSSPPPYYR

CPT_g0242 279975 279487

Polifobust homolog present in Genebank/EMBL as of 11/7/98

KSLKYCSLYQFSQKPTVILMACSIFFRMSQGDVDDEPLSKKTACLVVDTMLYPVIAVVCA

VGVVLLILKVLFLLLSFPFKLCSASSALPERVSLGSHFKCLYGGGLPYLLACLLIVPV

IGTAGHGFIISHRTSEDARLSSAIVFMQAPILQLAGMSGLIKP

CPA 1243 280609 280133 No Tobust homolog present in Genebank/EMBL as of 11/7/98 INNAMELVELLKFVKGRIIMACSIGYHLCNANEPDRFVASKVALVADILLYPFMAVICAVV FAYLMVVKLLFLAIKFLVNTCIAACKSRPLPSCKENFCCLFGPKDKPGPSDWLGCLVLIP IICTLIYSTIITVQSDTNRLRYFIISPAYQVGSTAIINW

280906 281556

CPHEU244
addrylate Kinase
GAFRATKGSVFIIMGPPGSGKGTQSQYLANRIGLPHISTGDLLRAIIREGTPMGLKAKAY
LDKGAFVPSDFVWEILKEKLQSQACSKGCIIDGFPRTLDQAHLLDSFLMDVHSNYTVIFL
EISEDEILKRVCSRFLCPSCSRIYNTSQCHTECPDCHVPLIRRSDDTPEIIKERLTKYQE

RTAPVIAYYDSLGKLCRVSSENKEDLVFEDILKCIYK

CPn_0245 281627 282499
ydho-Polysaccharide Hydrolase-Invasin Repeat Family
TCQKEIMKHYLSFSPSADFFSKQGAIETOVLFGERVLVKGSTCYAYSQLFHNELLWKPYP
GHSFRSTLVPCTPEFHIHPNVSVVSVDAFLDPWGIPLPFGTLLHVNSQNTVIFFKDILNH
MNTIWGSGTPQCDPRHLRRLNYNFFAELLIKDADLLLNFPYVWGGRSVHESLEKPGVDCS
GFINILYQAQGYNVPRNAADQYADCHWISSFENLPSGGLIFLYPKEEKRISHVMLKQDSS
TLIHASGGGKKVEYFILEQDGKFLDSTYLFFRNNQRGRAFFGIPRKRKAFL

CPn_0246 282955 282551 rs9-59 Ribosomal Protein WAKSTIQESVATGREKQAVSSVRLRPGSGKIDVNGKSFEDYFPLEIQRTTILSPLKKIT EDQSGYDLIIRVSGGGIQGQVIATRLGLARALLKENEENRQDLKSCGFLTRDPRKKERKK YGHKKARKSFQFSKR

CPn_0247 283430 282969
rll3-Ll3 Ribosomal Protein
DBY:IMEEKRUTKTTIVKSETTKSWYVVDAAGKTLGPLSSEVAKILRGKHKVTYTPHVA
WBJX:VIVINAEKVRLTGAKKOGKIYRYYTGYISGMPEIPFENMMARKPNYIIEHAIKGMM
PRTRLGKKQLKGLRIVKGDSYETFESQKPILLDI

CPn_0249 285850 284333 CP151 hypotherical protein

LEVMKFEFSVALKYLIPGRGRLYSAIVSLFSVGIISLVVWLSIVFISVIHGLEQRWIEDL SQLHSPITILPSDTYYSSYYYQIDKHSSLSNYTKTLGEKIASPQVDPYDPSDYLLPET FPLKDCDLCOQKDPVMMTLESLGPYLQSOHGKVIEFEQGVGYLDIKTSLKLQKPQPRNLTHFLTYPSKLSYEDKVLPYDETDYTSAELNPFNRSPSGWQDDFHHLEELYRGASIILPST YKDSGYKVDTGYFSTYSIENEKETQYTVHVIGFYNPGLSFLGGRTVFIDPDLARSIRSQ SEGLGMSNGFHLFFPNTKRIVFVKKQIENILTSLGYDDYWEISSLHDYDYFQFILDQLQS DQVLFLFYCILILIVACSNIVTMSMLLVNNKKEIGILKANGTSSRSLKIFFACCGAFSG ACGVVIGTIFAIITLKNLQFIVKALNYLQGRETFNTAFFQQNLPNSVHPQAIYFLGLGTL

CPn_0250 2#5069 L833702 r133-L33 Ribosomal Protein KDSSMASKNREIIKLKSSESSDMYWTVKNKRKTTGRLELKKYDRKLRRHVIFKEAR

286036 287559

CPn_0251 286036 287559

*conserved hypothetical protein

\$PDSCLPMMSPFKKIVNRLLCYISFQKESRTLPIIIREPRMTTKSLGSFNSVISKNKIHF
ISLGCSRNLVDSEVMLGILLKAGYESTNEIEDADYJLINTCAFLKSARDEAKDYLDHLID

VKKENAKIIVTGCMTSNHKDELKFMMSHIHYLLGSGDVEHILSAIESBESGEKISAKSYI

EMGEVPRQLSTPKHYAYLKVAEGCRKRCAFCIIPSIKGKLRSKPLDQILKEFRILVNKSV
KEIILIAQDLGDYGKDLSTDRSSQLESSLHELLKEFGDYWLRMLYLYPPEVSDGIIDLMQ

SNPKLLPYVDIPLOHINDRILKGMRRTTSREGILGFLEKLRAKVPQVYIRSSVIVGFPGE
TQEEFQELADFIGEGVIDNLGIFLYSQEANTPAAELPDQIPEKVKESRLKILSQIOKRNV

DKHNQKLIGEKIEAVIDNYHPETNLLLTARFYGQAPEVDPCIIVNEAKLVSHFGERCFIE

TYTTTAGYVDIJUGRVVKKSONOALLKTSKA ITGTAGYDLVGRVVKKSQNQALLKTSKA

CPn_0252 288112 287576
CT144 hypothetical protein (frame-shift with 0253?)
atstvCalwilQtYQShDDAASCSFRRACRFGRYWLGGVNVPWNKFNQTSTQSTVINSAI
YIDSSQTWMMRFQASASIPRLFRISIFHTKHGWIDNGTGGELLLVAYEANQNPLFPDIR IELAMSTCSGTSYYRARPMOWLCSTYYAVRPGYFVLENRSYSFRVQSFSWNIATLPFVN

CPn_0253 288474 287950
CT144 hypothetical protein (frame-shift with 0253?)
FCCGRLMSSSIPTTQKITISIPTFVRFNIESINLTDEQKKTALTIGQNIATEMTQVLGNF
VDADGGLICQNDLSVGGNINITPQTFNTMVFSGRVNLSNSFFSYQDSLGNKRQDYANINT
EQPQQYVPYGYYKLTRVMMMQRAALSGGHVGSGDIGWGESMYLGISSIKRQHKVQ

CPn_0254 289268 288459. CT143 hypothetical protein IPMKTLGVKDQNLFIDQATLSVERNVRIENNLETRDLKVLDTTTSPCEFIVKGNVSAEGS CHARTLESGENIYSKTDVSOTPVCNNISDPOSARDALTFSYYRKTGCOAANLYTYYPGN
GYYVAPNTIETHVAAITSKSVSRNATEDFSRYADIEPVVKLKQVGIYQVTMQLTRWSGQ
HDGDNSATLILNEVSGONKTLLCTSDTRGGYSSDRTSVAVTAIFSVTELVSSPPYDYPWI
NLESTIWMNLMSLSTCVIWFPFPSNFVEVD

CPn_0255 290183 289329

CPH_U255
CT142 hypothetical protein
TLLKVIMKNNINNNECYFKLDSTVDGDLLAANLKTFDTQAQGISSTETFSVQGNATFKDQ
VSATGLTSGTTYNLNAQNFTSSQISIDFKNNRLSNCALPKEDCDPVPANYVRSPEYFFCS
KPLIGDFDFNSGESYLPLTGSEYTLYQSRNVNSIFRFIGWKQSTRELTVGGNTAIQFLAA
GTYIVSFTVGKRWGMNGWGGAIYINNGLGQVQCESTIYSGGGYATIGTLGTSIYRASVD
VAPNPNDPNASDRYRAGIFYLSNGGSSAGIGNYSFSLLYYPDDRG

CPn_0256 291282 290398
CT144 hypothetical protein
FCGGRLMSNPTPKTKISIPTFVRFNIQSINLTEDQKKTTFTVGGKVTTENTVVRGDLTCT
DGGLTCQSDLTIQKDINIRPTSTNSMVFDGRLNLSNSPLSYKNSQGQDITDYEKMSSGKT
DGYVFGTYKRTQIMMAQRAAHSSGYVGGGSVPGSYVPWNKFDQTSTQKTSGTEIYIDP
NDSTKLVFEVNNKVPKLFRISVIMAKHGSWLDNGTGADILLAANEYEQGGGRINVTDLAM
TTSRGSSYYETRPLQVVCVTYYAQNNGYFTFQNRAGGGLRVSFFSWNIVALPYVE

CPn_0257 292136 291267
CT143 hypothetical protein
GVVMKRRNLQKILPNASTPSTNVAENTGIKDONLFLDQATLNVDGNVDIENFLETRDLKV
ADTITSPCEFTVGGGLSAESSQFKATTLSKGLEITSEDQDGRVPKFTNVSDPQSPRDALT
YNYYRNTGCQALNLYTYYSSSQPTTVGKPIETVCQNPNPETYRISASAKIYDAVTRFPYI
QFKARGIYQVTIQIRRESGGHSGLDNPNLYLNLMIGNNKTLLCASDTRGYSGGHRTSIAV TGTFTLTEIVATPPHDYPWLFLETTIGLDIKSMSTCVIWFPFQANFAEVD

CPn_0258 292534 292133
CT142 hypothetical protein (frame-shift with 0259?)
CFSFCRLGSKFEKITLGGNTAIQLLAAGTYILTFTIGKRWGWNNGWGGSIRLFEGKYTGD GTMLCGSTVYSGGGYSTIGYLSTAVYRDHSDIDPDPNNPSDKYMNNFLFVRNGDHSAVIG NYSFTLLYFAGDKV

CPn_0259 293031 292441
CT142 hypothetical protein (frame-shift with 0259?)
LYFVFKKKTYNYFIEMTTTNNQDNNECYFKLDSTVDGDLLASNIQTFDKQAKGISSTETF
SVQGNATFKEKVSATGLTSASTYKLNATGPAPSSITIDMKNNRLSNPALPKNPCDPVPAN
YVRSPQYFFCAKPIEGTFMFDGSSRYLPITGDGSNYTLYQSSKAGDVFRFVDWDQNSKKL
HLGGTQPYNFLLQEPIS

CPn_0260 294090 293548
secA-Protein Translocase Subunit
AYLDFSKRSCVEEDHVSKKINRIDLCPGSNKKYKQCCLKKEEQTARYTTEGKFKFSAEV
LSASEQGEAGDNCTKLFQRLSOSLTSEQKAAVGKFHQITKNKEVMSKKALKKAQAKEEKL
VTEKLQOHNFEILNTGENLAPPMESTATLNQDTNFVCEDFIPTQEDFRISENSOKPPVEE

CPn_0261 294272 295033
ydaO-PP-Loop Supertum:1y ATPase
YSEPHEP LYEMSTLLUNPEWMKACKR LESLVRKALYTHTMLANHRKIVVALGCCKDSUTL
LLMLKAL JORGE PELDEHAVNICOKYSOSAEVAKPYLTE ICOQLO I PERTI ESPYAPETP
ECYPOSQAPRRLIFOAAKE IGACA LAFYHHRDDLVQTALLMLLHKAEFACMLPVLDMVHF
GVTILRFLI I TPEEFWIRKFAKEI GFARYTORG PVVSIGCHAEQCIJKLLEEVF PLARHNIA
LA TORHGCSKSOM I LATOEHGGSKGOMI

295055 295913

CIE_UZAZ IMOTO ZONOTO BUEE-SUFE TUKE ACTO PRODPATANA LIPNINKEVKVYLVLMNKRIKITITNDDDITAKGMGCLVOADLEANIGDIYTAAPQAEQS

296174 297136

yqfU hypothetical protein STALSRRKLRVRPPSLAKYAFPGFRMSHGPRPTKFSFPLYFSKTLSWFILGGFLAACGVQ WCMATHERLIFOLD LIFE SHEELOOD FOR THE STANDARD STAND REFERENCE LESS OF ANTI-BARRAGER DAY, INDIPERCENCE, COMPATIVIONETS!

ENTERGES WITH MOTHER TO THE STRUKE AND ENGLISHED FROM THE STRUKE AND THE IIINKKKGYTYGQIILFYNFFIFALSGIVYKNWHTAFVSFLTYGIATKVMDMYILGLEDT KSVTIITSSPRKLGHILMETLGIGLTYIHAEGGYSGEPRNLLYVVVERLQLSQLKEIVHR **EDPSAFIAIENLHEVINGRRT**

CPn_0264 297730 297155
ubiD-Phenylacrylate Decarboxylase
MKRYVVGISGASGVILAVKLIKELVNAKHQVEVIISPSGRKTLYYELGCQSFDALFSEEN
LEYIHHSIQAIESSLASGSCPVEATIIIPCSMTTVAAISIGLADNLLRRVADVALKERR
PLILVPRETPLHTIHLENLLKLSKSGATIFPPMPMWYFKPQSVEDLENALVGKILAYLNI PSDLTKQWSNPE

CPn_0265 298632 297730
ubiA-Benzoate Octaphenyltransferase
KIIIVRLNYFLULVNFKYSIFSILFLSASTVFALSINEISQNLSFKEGFKISVFGAIAFV
FARTTGIVVNQCIDRFIDKKNTRTSKRVLPANLVSLNFAWVLSLFCSFLFLFLCKILRIF
SLGIASLTLMIYYPYMKRVTFFCHWGLGLVYTVAILMNFCAFAESGLSNRLCFLALLWGG
SVGMVIAANDIIYAIEDTEFDREEGLRSVPAHYGEKKAVEIAKVNLWVSYLAYIFSGFVG SLDKEFYFTAIIPLVVILKVVRMYSNYSKKDQEGESKFFLANIAIALSFLVSMTLFWSLS

CPn_0266 299181 299876
No robust homolog present in Genebank/EMBL as of 11/7/98
IMALDEINNONPSQQIASSTSQTSKINQDRKTFACTVTLLVVATLMILSGIVLLFTIGS
EGLSVPLSGILGTFAVTVGAVLFITGGLTILVRKSLGIEQKNEDLNFLKIKTPTPPARPLM
SKFSVTCSTTSIVLGMALLIGAVVSVFFLTGYLQLGLCAGLVGLGTALFVAGLARMSPRS LADQEGSGSADSQSNIVGIGEPKAAQEQKWYKMAVVRGEDGIPTAIRLTPEK

CPn_0267 300122 300910
No robust homolog present in Genebank/EMBL as of 11/7/98
VSIMSLNKTNALLNQPEPAVCLNAWDPKYINQDRKTFACTVTLLVIATLMILTTGVIVLL
AMAGSFGLSVLVSTIIGTSVTTLGTALFIIGLVKLIKKSLAWIQYQKYFQEVVKQKYEPFS IPKADNVHKLTSCLPSPLDIESPSPEASTPVSKLRIACSGVAIVLGVTLLIGAVVSVFFC
TCHQLALCVGFACLGTALFVGGLAGLRTHSLIAQGIMYLYLTYYLSSALEERNETVKDQ
RNEENTYLTEECRQQKREKALLE

CER_50268 300914 301318

NO robust homolog present in Genebank/EMBL as of 11/7/98

KOWALSLMSQCQSSSTSTWEWMKSFVPNWKNPTPPLSPIPSEDEFILAYEPFVLPKTDPE NAOANPPGTSTPNVENGIDDLNPLLGQPNEQNNANNPGTSGSNPTSLPAPERLPETEENS

302468 301476 CPRIDUCY

JULYOB

JULY

JULYOB

JULYOB

JULY

JU

CPE=0270 303343 302468
ywlq-SuA5 Superfamily-related Protein
SIFGWIVPDKKAQITFSLPEVMSAIHQGKIVALPTDTVYGFVLSLYASEAEERLYALKDR
EPSKAFALYVNSIEDIENISGYPLSPTAKKLAQLFPGAITLVVKHRNPRFPKETLAFRIV DHSYNREIVDHCGTLIGTSANLSEFPSALTAQEIFADFADHDLCIFDGPCSHGLESTWA SDPLYIYREGLISRSVIENIAGTEAKIFHRTSHAFSKHIKIYTVKNQEQLVSFLSGSLDF KGYKEEHPKPKNFYTRLREALKKKTPSIVFIYDINTSDYPELFPFLSPYYIE

303628 304362

305272

Lysophospholipase esterase

KLMTDYSFFRRKIGNIEAIECPGNPQDPIIILCHGYGSLADNLTFFPSICSFSKLRPTWI
FPNGILPLENDFRGSRACFPLNVLLLQELSRLYANGVGNLQEKYDELFDVDLETPKEALE
ELILNLNRPYNEIIIGGFSQGAILATHLVLTSQNPYAGALTFAGARLFNQGWEEGLKQCA
QVPFLQSHGYEDEILPYHLGAHLNDLLLTKLNGQFVSFHGGHEIPSVVFQKMQVTVPNWI
DPARG

304340

dnaz-DNA Pol III Gamma and Tau FNRQSDATYATWYMHLEEENQGWEALLRRVYHQEVPPAILLHGFTLPVLQDKAEQLASEI LLSSSPGSBHKVSQK.HPDIYQFPPEGKGRLHSIDLPRGIKKQIYISPEANYKIYIIHE ADRWTLAAISAFLKVFEEPPKHAVIILTTAKVQRLPRTIISRSLSIFIERGEKILCSKET FS'YLFR'YAQCETPYTEVSQIIKESSETDKQVLRDKVQRFMEVLLELYRDRYTLNLGLKAS ALMYPEHVKEILQLPLLPLDKVLLIVESACRSLNNSSSAASVLEWVAIQLVSLQYKEKEL VSVSPGQDLSN

CPn_0273 305853 305227

308368 GFT_0274

BYTA DNA GYCASE SUBURLE A

ECTTIMEDIKOE TVPKNLEEBMKESYLPYSMSVIIGPALPDIRKKLKPOORRVLYAMKOL

LICISAKHEKGAA ECGOTEGOYHEHGESVIYPTLVFMAQNMAMRYPLVDOGONFOSIDGO

FFAAMRYTEARLTIGAMYLMEDLOKDTVDIVPMYDETKHEPVYPPSKFFNLLCNGSGIA

VGMATHIPFINLUSLIEATLULLANPQASVDEILQVMFGPDFFTGGIICGSEGIRSAYTT

GROKIKVPARLHVEENEDKHRESIIITEMPYNVNKOPLIEQIANLVNEKTLAGISDVRDE

SUMUNITYVUEIKKGESSEIIINRLYKFTDVQVTFGAAMLALDKKLERTMSIHMTISAWI

RHRKEVIPPRTRYELNKAETRAHVLEGYLKALSCLGALVKTIRESGNKEHAKERIIESFG 305852

TTEPOALAILELRLYQUTGLEAEKIQKEYEELLNKIAYYKOVLSDEGLYPDIINELQDL LKHHKVARRTTIEFDADDIRDIEDIITNEJVIITISGDDYVKRMPVKVFKEQRGGRGVT GFDMKKGAGFLKAVYGAFTKDYLLIFTNFGQCYWLKVWQLPEGERRAKGKPIINFLEGIR PGEELAAILNIKNFDNAGFLFLATKRGVKKVSLDAFSNPRKKGIRALEIDEGDELIAAC HIVSDEEKWHLFTHLDMAVPFPHEKVRPMGRTARGVRGVSLKNEEDKVVSCQIVTENGSV LIVCDQGFGKRSLVEDFFETNRGGVGVRSILINERNGNVLGAIPVTDHDSILLMSSQQAIRINMQDVRVMGRSTQTVRLYHLKEGDALVSMEKLSSNENDDEVLSGSEEECSDTVSLR

310 744 , . . . R

FMDPREKNYLDASAITYLEGIQAVRERFUMYIGDTGITGLHHLVYEVVUHSIDEAMAGICS RIDVRILEDGGIVIVDNGRGIPIEVHERESAKQGREVSALEVVLTVLHAGGKFDKDSYKV SGGLHGVGVSCVNALSEKLVATVFKDKKCYQMEFSRGIPVTPLQYVSVSDRQGTEIVFYP DPKIFSTCTFDRSILMKRLRELAFLNRGITIVFEDDRDVSFDKVTFFYEGGIGSFVSYLN QNKESLFSEPIVICGTRVGDDGEIEFEAALGWNSGYSELVYSYANNIPTRGGGTHLTGFS TALTRVINTYIKAHNLAKNKLALTGEDIREGLTAVISVKVPNOFDEGOTKGKLGNSDVS SVAQOVVGEALTIFFEENPGIARMIVDKVFVAAQAREAAKKARELTLRKSALDSARLPGK SVAQVVCBALTIFFEENVIARMITUNFVAAARARARARALITARSALSARAFGA LIDCLEKDPEKCEMYIVEGDSAGGSAKGGRDRFGAILPIRGKILNVEKARLOKIFGNQE IGTIIAALGCGIGADNFNLSKLRYRRIIIMTDADVDGSHIRTLLLTFFYRHMTALIENEC VYIAQPELYKVSKKKDFRYILSEKEMDSYLLMLGTNESSILFKSTERELRGEALESFINV ULDVESFINTLEKKAIPFSEFLEMYKEGIGYPLYYLAPATGMQGGRYLYSDEEKEEALAQ EETHKFKIIELYKVAVFVDIQNQLKEYGLDISSYLIPQKNEIVIGNEDSPSCNYSCYTLE EVINYLKNLGRRGIEIQRYKGLGEMNADQLWDTTMNPEQRTLIHVSLKDAVEADHIFTML MGEEVPPRREFIESHALSIRINNLDI

311140 CPn 0276 310793 CT191 hypothetical protein
DMFLKRKKRGGSQVQNKRTASPIKHAKHYLHNYLQELQKIMAARPHDAIDAWNQVFRDKY KGMSQAIGFRDHILLVKVYNSSLYALLKQTPQNDLIMSLYQVASHVQIREIQFLLG

CPn_0277 312003 311404 No robust homolog present in Genebank/EMBL as of 11/7/98 NISIFYPRYFIEGKEVLIKNIPPLIFYGVILMIINVRAPAFGITSVOOFSTNFQAAIPIL NIVIGCSRISSTYAEDIEEVAQEKLEKSTHSKSSTSVNLWAHRVRGVVEILGGGIVILAL EITALVLQVIIKLIKCLIDVLCVCLFGLGVCVVAIIGAIAFCVVVVVKYLGFCSQGEELE PIEVKTLISPDKPYPTVVYV

CPn_0278 312884 312060
*conserved outer membrane lipoprotein
RDSMKKKLSLLVGLIFVLSSCHKEDAONKIRIVASFTPHAELLESLQEEAKDLGIKLKIL
FVDDYRIPNRLLLDKQVDANYFQHQAFLDDECERYDCKGELVVIAKVHLEPQAIYSKKHS
SLERLKSQKKLTIAIPVDRTNAQRALHLLEECGLIVCKGPANLNMTAKDVCGKENRSINI
LEVSAPLLVGSLPDVDAAVIPGNFAIAANLSPKKDSLCLEDLSVSKYTNLVVIRSEDVGS PKMIKLQKLFQSPSVQHFFDTKYHGNILTMTQDNG

CPn_0279 313546 312875

* Possible ABC Transporter Permease Protein

KKINGSDLIQILLKETVNILYMVSTAFFFSCAIGGMLGLGLFCTSPKSLNPKKSLYATIS

MILSFLTAIPFAILIVILFFITRMIVOTSLGFTASIVPLTIGAIPFVVTIVVDAFRNSAL

NYLESAVALGIPKRNILFGILLPESYPQLIFSLKSLVVHLISCSTLAGFVGGGGLGQLLL QYGYYRFEWSVTTSVLVITLVLIESVRILGDFWGRRVLKYRGIL

CPn_0280 314593 313550
dppF-Dipeptide Transporter ATPase
IKGEAWLVSEOHSPIISVODVSKKLGCHILLSKVSFSVYPGEVFGIVGHSGSGKTTLLRC
LDFLDMPTSGSISVAGFDNSLPTOKFSRRNFSKKVAYISQNYGLFSSKTVFENIAYPLRI HHSEMSKSEVEEQVYDTLNFLNLYHRHDAYPONLSGQQKQKVAIARAIVCQPEVVLCDEI TSALDPKSTENIIERLLQLNQERGITLVLVSHEIDVVKKICSHVLVMHQGAVEELGTTEE LFLNSENSITNELFHEDINIAALSSCYFAEDREEVLRLNFSKELAIQGIISKVIQTGLVS INILSGNINLFRKSPMGFLIIVLEGEVEQRKKAKELLIELGVVIKEFY

CPn_0281 315033 316103 "dhnA-Predicted 1,6-Fructose Biphosphate Aldolase (dehydrin family) * ISLRRHTLMLNIHDILGNDDENLLSYQCKHITKDKLTLPSHDFVDKVFGLSDRNNRVLRS LOTMFSHGRLANSGYLSILPVDQGIEHSAGASFAINPIYFDPENIVKLAIESGCSAVAST YGTLSLLSRKYAHKIPFMLKLNHNELLSYPTKYHQIFFTQVEAAYSMGAVAVGATVYFGS ETSNEEIVAVSNAFAKARSLGLATVLWCYLRNPAFVANGVDYHTAADLTGQADHLGATLG ADIVKOKLPTCOGGFKAINFGKTDERVYSELSSNHPIDLCRYQVLNSYCGKVGLINSGGP SGKNDFTEAARTAVINKRAGGMGLILGRKAFORPLSEGIOLLNLVODIYLDPNITIA

CPn_0282 316084 317529

xasA/gadC-Amino Acid Transporter

ILILOSLNFSKRVFMHSHSKPTKPLGTETVGMLSLAVVISLRNLPLTAKHGLSTLFFYGL
AVICFMIPYALISAELASFKPGGIYIWARDALGKWWGFFAIWMOWFHNMTWYPAVLAFIA
STIVYKINPELAHNKYJATVILAGFWILTFFNFLGITSSALFSSICVIIGTLIPGVILV
SLALFWIFSGNPIAISLSWGNLLPNFSNVSSLVLLAGMLLALCGLEANANLASDMYNPRK
NYPKAVFIGAIATLTILVLGSLSIAIVIPKEEISLVSGLYKTFTLFFDKYNLSWMTGIVV
VMTIAGSLGELNAWMFAGTKGLFISTONDCLPRLFKKVISKNVPTNLMLFQGIVVTIFTL
LFLCLDSADLVYWILTALSVQMYLAMYICLFLAGPILRIKEPRAQRLYSVPGKFLGICTM
SILGILSGAFALWSSLPPRELAGDISGGSK IGVTTPLILAFSLNCLIPFGIVFTHKRLSK SILGILSCAFALWVSFLPPRELAQISEGSKIGYTTFLLLAFSLNCLIPFGIYFTHKRLSK

318581 CPn_0283 318581 317532
No robust homolog present in Genebank/EMBL as of 11/7/98
GRRLSYFODLIKNAVAKIISFRKSPPNPVKLLIKFAKKDLKNSSIAPLYEVLLEILEAPG
EEILEVLFSLDPMWLKSMLDPKKHSTLGIEISSETAETIESCSLGLISINLLLSGLCLRS
SHDROQAVKIIQOFCPGFSSEEVONFVEORNILTPFLHHLFEODEVALLNQLSLRLDLIV
PNALYZEPPDSCWOSINSECAKDADEOQEDFNKTKEACYEGLKKLVLPALSITSIFQLIV
RARRFKQGAEILMAIDRKKMKQNPFIFLEALLEGDEFGIOVGKYLKLLMPIHLWDKLLHA
LYLGYFOYDLICOGELETFGRRANLNPEAFQAAIQQOFLLGFLFFKMLLD 317532

CPn_0244 4 49054 318551
No robust homolos procent in Genebank/EMBL a., of II/7/98
PLLMMIPARQVPVIDEPVVNNTSSWILSEKSELPPTTVLJEATLATEATATEMSVLYFCGTTS
VOTEVIJAMLIPESVCSVLCVAYLEYQQGSTEPTTVPGTTTSPSVETSDEDENILJEREEDS VEATDELLKNEPADDERREKMLEYONFLDEGGEPRESPEEDSHTGKIL

320601 319051 CHILDON SERVICE STORY OF THE SERVICE STORY OF CONTROL OF THE SERVICE STORY OF SERVICE STORY AAPLGLLVWGCAASVCSMMAIVSLMCLYKGGKPLIEPSNEEKIDPTKDLEIKDPESLKPV PVEGGSLPKERKTVSFKAKIPSIVEDDFKPYVIGSTFYHQNKVYSKPIAERMGSLEKEIT TLIVDFPRALEESSK3SGSLLRGVISEIKNLFLPRFLSRKVKYSLTACLPRLGSIVEEYA SSDLLILLLTKPEPLMVTQQLIAHLNSLKTEKRKLTPHMOKLVLSINFWFYGWSLEEKC IEKIVAYDPNLLTDELKAHLEAGNIVQFLLSFGSSEMQREFRALFPSDAQELPSAKDGSN YVPAINSSEYMTOFKDLSVLKKSLSERLAFCEKIPSPSSWNFTSSVASHYKDFSLLFTFF SNQGSVILQNPFLLIELLHENPKCCTFLKGLLEKAMPMSWAALFRPMLMGMLCSGIARK KELKIIAEHLGVPFKEITQAIASGKILDLLLQHLFDF

MgtB-Mg++ Transporter (CBS Domain)
SCRESKGKIMVGEQNRNEEKLDTAFSSGNIMDSRTSHLDDELSFKLEKAFTCLSTDIHSH

SCRESKKIMVGEONRNEEKLDTAFSSGNIMDSRTSHLDDELSFKLEKAFTCLSTDIHSH
DLSKIVIEYNPIDLAYAVSCLPSESRAILYKNLSCITAKVAFIINTDSASRWAIFRRLSD
SEVCALIEOMPPDEAVWYLDDIPDRRYRRILELIDSKKALKIRDLOKHGRNTAGRLMTNE
FFAFIMETTVKDVSACIRSNEGIDLTRLVFVLDFKGELOGVVTDRSLIINPPEMSLKOIM
NOIEHKVLPDATREEVVDLVERYKIAALPVVDEENFLIGAITYEDVVEAIEDIADETIA
MAGTTEDVGYOTCHVVORFLLRAPWILVTLFAGLISASVMAYFOKISPALLALIIFFIPL
INGMSGNVGVQCSTILVRSMATGTLSFGRRRETIFKEMSIGLLTGVVLGILCGLVVYLMG
FLGLNIFSGGGIOLGVTVATGVLGASLTATTLGVLSPFFFAKLGVDPALASGPIVTALND
IMSMITERILAGGINDFI.PPN IMSMIIFFLIAGGINFLFFN

CPn_0287 324230 322089

No robust homolog present in Genebank/EMBL as of 11/7/98

RRCMIRSPLFFISSKRALNMLGLQDEFSCFEDVVDFLFSEIELLAQODEPSEGYLALSRS

LLMMTHNHPKVVKKVIFYGVSYGLKHKSMSIFIDVLTYIDFLFEKLGISASDRLSLGSAR

TCINFELYSQTGEMKFLSEVVDNFFLIEQLLKHPQLKNRLGWEHFRIGAKQEEVSLVAS

ASVYQAVGRSFIELYHKHLELSDLACGMKCLALALDLSPNNAHIHADYAKGLVVLGTRQS

KSLLIERGMEHFSKA IFLSFSRGEDTLAYONYRYSYALASVKLFGLTYKKEHFDQAMNIL

YQTVQAFPNLSGLWMVWGELLIRSGWLNSNMKYIEVGLEKLASLQKKTNDPIALSGLLAT

GIAILGLYLEEDNLFKDSRHRLISAMRTFFGNSALVHALGVVQLCSALYFNEDSHFASAI

SCFOSCLEDDLDATGMOKKLEDAYSWGIKKSARLLRKAUTVASRICSLRFBAFLFWSD SCFOSCLEWDLDATGMWOKLFDAYTSWGIKKKSARLLRKAVDVASRLSCLEREAFIFWSD
RGLALKCLAEATIDEAYKEIFLSESLLHYORAWDLSGRLEILEILMGOSHYLLAELQOSLF
HYDEAYTLLTKVDLTLSSRVKLILAAVLLGKGRLLQDTDPAEEAREILEPLVEVYLEDE
NFLLLLGKVYLFLFWKNKNVCLGKLARTYLEKATSLGCPEAYYTLGKFYAVIKDVNRAWG MVIRSAQYGVRITEAKWLNDPYLANLREIHAFREVVENQKGRLWLGNKTEMKRN

CPn_0288 325785 324571
CT288 hypothetical protein
ISITIREFLEFGECRAKFYNVIMSCFRLTSTNESLRPISPKASFPKQGWQSYFRSALRK
HRSDTLSVSVCKVNKYDANLFVRLTVIALAVVGVLILFSIMLASIQGTLVITSWPLVTAA
IÉLPTILLTGGMYILHRIGKKVDVISGVCIPPFSRRCMVPISSSHTLEKFDEKHVSACSY
LDISTLSADGSGIAAVYQCPPLLFRAFPCFGIPCAMPFVALLRMIYNLIRFLVVPFYIIF
RMIYEHFFCKHLDEDDRFIYKDVAREMGRSLAAFLKAPFYASACMIGAFYSLLDPLAGRV LMGSVERDWNDNVILARSVSLANEAHSLFRFEGGGGRKGLGQHAFYLMLCCQPQSVFLFD KGETVSGAHPSIQLPERGLDTSGRYPHISVIPDSGNDSAKNFIV

325797 326996

CPh_0289 325797 326996

CT239 hypothetical protein
nenrimkkorshykknnllllisilvGlgGgsvQspwivysaeciantflkflrllsipl
veeligstitsiQnfntmvtigkrilyytllttviaasiGlllffllromitQdalatt
tkcnplgyldvlsdtlpenifkpflQcnvisaaclavllgtaslfQekekhfvNQffns
fristflnlargGlklpiaMlgfsvilfkelktQsnltmfaeytlcviGanlaQgftvlp
iqtkinkvsplkvakamspalvtaffskssaatlpltmelaeddlkinknlskfsfplcs
vennkgCaafilitvlfvatsngmiisplmslgwifiatlaaignagvpmgcyfltlsll
tsmwvplstiglit.pfytvtibmietslnvwsDccvvslan TSMNVPLSILGLILPFYTVIDMIETSLNVWSDCCVVSLAN

CEn €0290 327027 328523

CRALO290 327027 328523
Nä-dependent Transporter
RSALTMINKKHASFSSRLGFIFSMIGIAVGAGNIWRFPRVAAQNGGGAFLILWLCFLFLWS
IËLIIIELSIGKLTKKAPIGALIKTAGKKFAWAGGFITLVTTCILAYYSTIVOWGLSYFY
YÄVÄGKIHLGNDFAKLWTSHYQSSIPLWAHLTSLGLAYLVIRKGIVHGIEKCNKILIPAF
FLCTIALLLRAVTLPGAVQGIKQLFSCDKSCFSNYKWIEALTONAWDTGAGWGLLLVYA
GFREKKTGVVSNGALTAICNNLVSLIMGIIIFSTCASLDILGTTQLQDGAGASSIGITFI
YLPELFTRLPGGIYLTTLFSSIFFLAFSMAALSSMISMLFLLSQTLAEFGIRPYISETLA
TÜÄEVLALIFSANGTIFFSNQDTVWGVALIVNGLIFIYAALVYGFPKLKKEVINAAPGDL
RLNKAFDYIIKYLLPIEGILLLGWYFYEGLFPENGQWWNPISLYSLGSLVLQWSLGLIIL
WKENKQLYLRFSRYNHEIL

CPn_0291 328658 329194 incB-Inclusion Membrane Protein B EKHMSAPIPTPQELSDQITCLNVQYQOVSELARENKGDIEGLKTLTAALTADAGIQPSAD EIYSLQTAAALILSASEKPGSGPSGSTEGSVTVQSPCKFKKVLAVVLTIIALIAIAVLIA CIIAACGGFPLLLSALNLYTIGACVSLPIIASTSVALICLCTFVANSLIKPVITVRTTR

CPn_0292 329201 329836
incC-Inclusion Membrane Protein C
VKNTKNSDFMTSPIPFQSSGDASFLAEQPQQLPSTSESQLVTQLLTMMKHTQALSETVLQ
QQRDRLP7ASILQVGGAPTGGAGAPFQPPADDHHHPIPPPVVPAQIETEITTIRSELQ
LMRSTLQQSTKGARTGVLVVTAILMTISLLAIIIIILAVLGFTGVLPQVALLMQGETNLI
WAMVSGSIICFIALIGTLGLILTNKNTPLPAS

CPn_0293 329940 332723
CT234 hypothelical protein
WSMGRVTRLIFNLHHGEBRRAFLFFLLGLVWGIGCYCTLSLAEGLFIEKLGSAELPKIY
UGSSLILCVISSLILVNLFKKHISATALFLIPVSLSILCNFYLILSSIFAIDPPRSPLFF
YRIVIWSLTILSYTSFWGFVDQFFNLQDGKRHFCIFNAIIFLGDAIGSGIIASLVHTIGI
OGILILFTAALUTFFIVFVYSKSLKSLSDDHDLFIDTGHPPPLSKALKLCFYDKYTFYL
CFYFLMQLLAIATFFNYLKIFEIDFASKEEFELVAHIGKSLMISLGMMCFALFAYSRI
VKRLGVNNIILFAPLCFLSSLFLFWFFKTTLSIAVLAMVVRGGVTYALDDNNLQLLIYGVP
NKIRNQIRIVVESFIEDIGMLVWSLICFLGSQQVVFCLIISLIATILVCLVRGYYAKAIL
KNLGAQALQLTRSMQDWIKGMTVKQKROVELFLLAHLKHPSERHOTFAFOHLINLASRSV
LPGLLAHMNKLSLPNRLKTIEMVKSSLMAKDFLTLELLKRWTSIFPHFAIACAIHLYFAE
HDLLHITHIAEDLYDTVGDRLLAA LITVRRQEGYYGPYRDLADKRLKELLNSDOPEDIVMG
LTILKLEKNPQNFPILLDFLWTKNEDILIVTTCALHTGVRANHKPYCFELLFRLROCSHN
DEAGOYLLKTISIALDISFVKDLLMTTDQLKNTGRKYAGAMIGELDKEVAPAFLOVLTDE OTHNROR FLAAKALOK FUNWELKKHAYK EVKSKASKAEFYSYHOHY EQKKYPTYNESELA NTEMBNYYAFYNFMESELG FLOSMEHSOVE FRAETSKNOK FKAQALESEEKHYDSHEFSE LEPFVNOFYMCYSEKYYFKCOV FPETEKEELINMMENSPOSENKUTAQQEKEELSYCDPDF QSVNTIYNQEHEDFRTEESETLISFLSI

CPn_0294 333077 133502 CAMP-Dependent Protein Kinase Regularory Subunit IRNEFMNLIDRAFLLKKTIIFQSLDMDLLLTIADKTETLIFKPGSNVFSIGQPGFSFYII VEGYITISKEKLESPLNLKPLDCFGEESLFNNKPREYNASANTQVRMLVLSKGQILNIVE ECPSVALSFLELYAKQIKFREP

333627

acpP-Acyl Carrier Protein
AMSLEDDVIAIIVEQLGVDPKEVNENSSFIEDLNADSLDLTELIMTLEEKFAFEISEEDA

334732 224022 CPn 0296

CT196 hypothetical protein
KIPIRGMICMDITLYGKKVIVTGGSRGIGLGIVKLFLENGADVEIWGLNEERGQAVIESL TGLGGEVSFARVDVSHNGGVKDCVQKFLDKHNKIDILVNNAGITRDNLLMRMSEDDMQSV ISTNLTSLYYTCSSVIRHMIKARSGSIINVASIVAKIGSAGQTNYAAAKAGIIAFTKSLA KEVAARNIRVNCLAPGFIETDMTSVLNDNLKAEWLKSIPLGRAGTPEDVARVALFLASQL SSYMTAQTLVVDGGLTY

CPn_0297 335724 334774
fabD-Malonyl Acyl Carrier Transcyclase
SRSNKDDNFMKRRYAFLFF0GGSQYVGMGQDLYMEYPEVRELFDFANERLGFSLTSIMFE
GPEDLLMETVHSQLAIYLHSMAVVKVLSQRSSIQPSLVSGLSLGEYTALVASDRISVLDG
LELVRKRQQLMNEACNQSF0AMAALLGLPSEVIEENITSLGQGIWIANYNAFKQLVVAGI
AEKVDQAIELFRDLGCKKAVRLKVSGAFHTPLMCVAQDGLAPDIYALCMKDSSLPLVSHV VGKSLVNTEEMRECLAROMTSPTLWYQSCYHIESEVDEFLELGPGKVLAGLNRSIGISKP ITSLGTFAQIEKFLSEV

336742 335717

CPn_0298 336742 335717
fabH-Oxoacyl Carrier Protein Synthase III
YTSFFLYMMFSVNKNKAAIWATGSYLPEKVLSNADLEKMYDTSDEWIVTRTGIKERRIA
GPQEYTSLMGAIAAEKAIANAGLSKDQIDCIIFSTAAPDYIFPSSGALAQAHLGIEDVPT
FDCQAACTGYLYGLSVAKAYVESGTYNVLLIAADKLSSFVDYTDRNTCVLFGDGGAACY
IGESRPGSLEINRLSIADOKLGELLSLPAGGSRCPASKETLQSGKHFIAMEGKEVFKHA
VRRMETAAKHSIALAGIQEEDIDMFVPHQANERIIDALAKRFEIDESRVFKSVHKYGNTA
ASSVGIALDELVHTESIKLDDYLLLVAFGGGLSWGAVVLKQV

336726 337415

CPIL 1277
recR-Recombination Protein
RKKLYYYSESLYSNLANGPRPECKNK HIITMTRYPDYLSKLIFFLRKLPGIGFKTAEKLA
FELISWDSEQLKILGNAFHNVASERSHCPLCFTLKESKEADCHFCREERDNQSLCIVASP
KDVFFLERSKVFKGRYHVLGSLLSPITGKHIENERLSILKSRIFTLCPKEIILAIDATLE GDATALFLKQELQHFSVNISRLALGLPIGLSFDYVDSGTLARAFSGRHSY

340152 337768 CPn 0300

yaeT-Omp85 Analog GRLLGMLIMRNKVILQISILALIQTPLTLFSTEKVKEGHVVVDSITIITEGENASNKHPL GRLLGMLIMRNRVILQISILALIQTPLTLFSTEKVEGHVVVDSTTIITEGENASNKHPL
PKLKTRSGALFSQLDFDEDLRILAKEYDSVEPKVEFSEGKTNIALHLIAKPSIRNIHISG
NQVVPEHKILKTLQIYRNDLFEREKFLKGLDDLRTYYLKRGYFASSVDYSLEHNQEKGHI
DVLIKINEGPCGKIKQLTFSGISRSEKSDIQEFIQTKQHSTTTSWFTGAGLYHPDIVEQD
SLAITNYLHNNGYADAIVNSHYDLDDKGNILLYMDIDRGSRYTLGHVHIQGFEVLPKRLI
EKQSQVGPNDLYCPDKIWDGAHKIKQTYAKYGYINTNVDVLFIPHATRPIYDVTYEVSEG
SPYKVGLIKITGNTHTKSDVILHETSLFPGDTFNRLKLEDTEQRLRNTGYFQSVSYYTVR
SQLDPMGNADQYRDIFVEVKETTTGNLGLFLGFSSLDNLFGGIELSESNFDLFGARNIFS SQLDPMGMALYKDIFVEVKETTIGNIELFILEFSSLDNETGGIELSESANFGARRIFS KGFRCLRGGGEHLFIKANFGDKVTDYTLKWTKPHFLNTPWILGIELDKSINRALSKDYAV QTYGGNVSTTYILNEHLKYGLFYRGSQTSLHEKRKFLLGPNIDSNKGFVSAAGVNLMYDS VDSPRTPTTGIRGGVTFEVSGLGGTYHFTKLSLNSSIYRKLTRKGILKIKGEAQFIKPYS NTTAEGVEVSERFFIGGETTVRGYKSFIIGPKYSATEPOGGLSSLLISEEFQVPLIRQPN ISAFYFLDSGFVGLQEYKISLKDLRSSAGFGLRFDVMNNVPVMLGFGWPFRPTETLNGEK

CPn_0301 340163 340762
(OmpH-Like Outer Membrane Protein)
IKDLSKEIFVVFRKGFWYPFSIPKLVQVIMKKLLFSTFLLVLGSTSAAHANLGYVNLKRC
LEESDLGKKETEELEAMKQOFVKNAEKIEEELTSIYNKLQDEDYMESLSDSASEELRKKF
EDLSGEYMAYQSQYYQSINQSNVKRIQKLIQEVKIAAESVRSKEKLEAILNEEAVLAIAP
GTDKTTEIIAILNESFKKQN

CPn_0302 340766 341866

1pxD-UDP Glucosamine N-Acyltransferase
SKFKEFSMSEAPVYTLKQLAELLQVEVGONIETPISGVEDISQAQPHHIAFLDNEKYSSF
LKNTKAGAI ILSRSQAMQHAHLKKRFLITNESFSLTFOKCIELFIEPVTSGFFGIHPTAV
IHPTARIEKNVTIEPYVVISOHAHIGSDTYIGAGSVIGAHSVLGANCLIHPKVVIRERVL
MGNRVVVQPGAVLGSCGFGYITNAFGHKKPLKHLGYVIVGDDVEIGANTTIDRGRFKNTV
HEGTKIDNQVQVAHHVEIGKHSIIVAQAGIAGSTKIGEHVIIGAGTGITGHISIADHVI
MIAQTGVTKSITSPGIYGGAPARPYQETHRLIAKIRNLPKTEERLSKLEKQVRDLSTPSL AEIPSEI

342982 341921

CT103 hypothetical protein
REQKGLHHMDVSRKINRHTQFYVDSIDGVIKNFDHKPSEDKSRDHEELEEKLLTITKRIV ASAQEFONRKTDSKNYYLKKTOWLPFKNEELEOTKELFAMLTSMDKKTAQLFFYSPGCSS DWVEFTEVICHLNDSIGLGGVLLCCGLFEQQCEHVVTVNKKLDLPLLLGTTVVNSLRYYL TYRNISLLNCQSMSELGKELGDVLKQHGVAFTLIFKEIVDIDLLNYVKLIQGLKRSGNIQ ARIYDNDVPTLPSVSSSPIALRYSLANTIRGLALHVDFSSLKFISPSILSNTEHTAKALN GGGECFIFSNLDEFNLGMKIVMQLLRTGKISPEILNKNIMKILMIKRRVRSLYI

343091

DQKPLPKKLPYKKVMDSSAPYNIADGYTEKCTVERILDLYGPASCIKFLKQMVLIREFEA
RGEEAYLEGLVGGFYHSYAGQEAVATAAIANTGLDPWVFSSYPCHALAILLNIPLQEIAA ELLCKETTY ALTROGOMENC GENEFEYGEG LVVVO I ELAAGAAFT I KYOEOKIRVSLEFIG DGAVAÇ VEHTTLINTYOLIQLELMLI I ENNOWOMOTSLINRAVAKQE I AEDGOSSYD TRAV TYNGE OLD II LLGEFREAYRYMVDUEGDVI.VIV I LOGEFREGE I ODEN LYGGKEIMOCLEKK DE IVLAKUML I RLEVITTE EFON I ROECKTAVI CAAFGNAKLOODEN VITTLEECVYA

345137 344142

EPEANTROEALLEEATAL FORMHANOLUMPER LILAGUMAGILIRRACH VRRAMFDIET ABUNDANYANGAGORIBRAGARDIAN LILUT I TANGMIADAKATRAT HUMBARTERN EPEREWARGIG LOVATSIENET EELMISMILGILANTRO FORMIR MILKIREGAAL BRIEDAGORIBRATER DORMAGINA TINGERAGDARAY KALKOTTORMIBRIA CBU 3339, 44148, 34243,

[DLRT[KF:LD]GTILSGVRKTGRC[V]EEGHYFAG]GSE]IALITEHVFDSLDAPPLRVCQKETPMPY::KILEQATLPNVNRILDTIEKVMR

346431 345136 CPR_U306
pdhC-Dihydrolipoamide Acetyltransferase
GKFVISLLKMPKLSPTMEVGTIVKWHKKSNDQVSFGDVIVEISTDKAILEHTANEDGWIR GREY VISLIKHENDER HEVETTEVANTENDEN OFFINDEN FOR THE ANEXCHI AN VAIPDGIITPIIRCADRKNLUMISAEIKSLALKARNQSLQDTEYKUUSFCVSNLUMTGIT EFTAIVNPPQAAILAVGSVTEQALVLDGEITIGSTCNLTLSVDHRVIDGYPAAMFMKRLQ KILEAPAVLLLN

CPn_0307 348999 346515
glgP-Glycogen Phosphorylase
NGCIVEDFSSFDKNKVSVDSMKRAILDRLYLSVVQSPESASPRDIFTAVAKTVMEWLAKG
WLKTONGYVKNDVKRVYLSMEFLLGRSLKSNLLNLGILDLVRKALKTLNYDFDHLVEME
SDAGLGNGGLGRLAACYLDSMATLAVPAYGYGIRYDYGIFDQRIVNGYQEEAPDEWLRYG SDAGIGNGGIGRLAACYLDSMATLAVPAYGYGIRYDYGIFDORIVNGYQEEAPDEWLRYG
NPWEICRGEYLYPVRFYGRVIHYTDSRGKQVADLVDTQEVLAMAYDIPIPGYGNDTVNSL
RLWQAQSPRGFEFSYFNHGNYIQAIEDIALIENISRVLYPNDSITEGQELRLKQEYFLVS
ATIQDIRRYTKFHICLDNLADKVVVQLNDTHPALGIAEMMHILVDREELPWDKAWEMTT
VIFNYTNHTILPEALERWPLDLFSKLLPRHLEIIYEINSRWLEKVGSRYPKNDDKRRSLS
IVEEGYQKRINMANLAVVGSAKVNGVSSFHSQLIKDTLFKEFYEFF PEFFINTNTGYTPR
RWIALCNPRLSKLLNETIGDRYIIDLSHLSLTSFAEDSGFRDHWKGVKLKNKQDLTSRI
YNEVGEIVDPNSLFDCHIKRTHEYKRQLNNILRVIYYYNDLKENPNQDVVPTTVIFSGKA APGYVMAKLIIKLINSVADVVNQDSRVNDKLKVLFLPNYRVSMAEHIIPGTDLSEQISTA GMEASGTGNMKFALNGALTIGTMDGANIEMAEHIGKENMFIFGLLEEQIVQLRREYCPQT ICDKNPKIRQVLDLLEQGFFNSNDKDLFKPIVHRLLHEGDPFFVLADLESYIAAHENVNK LFKEPDSWTKISIYNTAGMGFFSSDRAIQDYARDIWHVPTKSCSGEGN

349213 349596

No robust homolog present in Genebank/EMBL as of 11/7/98 FFTQENNMATVAQTPQTTQPQPSVSHKATHRYCSWVFFKPILVSLGLLLASLTTLGLVIA SGVTLSLGIGIVLAIQIVLAGIALVLAFNHIRQFKQARTAELNSMKMISAPAAATVQKQK

CPn_0309 350977 349595
CT309 hypothetical protein
FMRAWEEFLLLOEKEIGTNITUDKWLRSLKVLCFDACNLYLEAQDSFQITWFEEHIRHKVK
SCLMNNNNKPIRVHVTSVDKAAPFYKEKQMQQEKTAYFTMHYGSVNPEMTFSNFLVTPEN
DEPFRVLOEFTKSPDENGGVIFNPIYLFGPEGSGKTHLMQSAISVLRESGGKILYVSSDL
FTEHLVSAIRSGEMQKFRSFYRNIDALFIEDIEVFSGKSATQEEFFHTFNSLHSEGKLIV
VSSSYAPVDLVAVEDRLISRFEWGVAIPIHPLVQEGLRSFLMRQVERLSIRIQETALDFI
IYAISSNVKTLLHALNLLAKRVMYKKLSHQLLYEDDVKTLLKDVLEAAGSVRLITPLKIIR
NVÄGYYGVSQESILGRSQSREYVLPRQVAMYFCRQKLSLSYVRIGDVFSRDHSTVISSIR
LEDGKIEENSHDIHMAIQDISKNLNSLHKSLEFFPSEEMII

CEN_U310 353472 351049
601M-60kDa Inner Membrane Protein
YFDELSLIFRVYQMNKRTLLFVSLIGIAFVGCQIFFGYNEFRSCKNLAEKQRKISEQTLA
AVESVGLSVASWDTDVNGEEHKNNYAVRVGDKLFLLENGEAAQSVYSSGESWSFVDHKCG
FDNHHLALVRQQGSFSPFNTNTGKVRLPTHNEGLPVLVVEFRNNKEPLVFLGEYAQGRISN
KDSTIFGTALVFWRSGSDYIPLGLYDGSREEKLVSLDLPITRAVIFGNDQDSAKSSDTANH
YVJFKDYMQIIVSEESGSIBGINLFFASTNNKSIVNEIGFFDRDLASEKSPEALFRGLSSK
KDSTIFGTALVFWRSGSDYIPLGLYDGSREEKLVSLDLPITRAVIFGNDQDSAKSSDTANH
YVJFKDYMQIIVSEESGSIBGINLFFASTNNKSIVNEIGFFDRDLASEKSPEALFRGLSSK
KDSTIFGTALVFWRSGSDYIPLLRGGLISDSKKLLPLEYHALNVYGGRELAFPVALRYRVLSYTP
HSTQLESLDRSVQKVYKLPENPEERPYVFETATTLTKETEDVWYTSGVPEVEIMSNASAP
TIKGKFVIKNKNGSSLDKVXLPRVKEPLAIRRGVYPQWILNSNGYFGIILTPLSEIASGYGS
LYTSGSTAPTRLSAISPKNQLYPVSKYPGYTETLPLPLKDAGTHRFLVYAGPLAEPTLKVL
DKT##TQEKGENPEYLDSISFRGVFAFITAPFAALLFIIMKFFKLVTGSWGISIILLTVFL
KLLYPLHAWSIISRMRMQILSPYTQQIQGVKNEPEKRAGMEIMGLYKTNKVNPTTGGLP
LLTGLPFLIAMFDLKSSFLLRGASFIPGWIDNLTAPDVLFSWQTSIWFIGNEFHLLPIL
LGIVMFLQQKVTSLHKKGPVTDQXKQQVMGNMMAILFTAMFYNFPSGLNIYWLSSMILG
VVQONITNKILDSKHLKNEVVLNNKKHR

CPh_0311

CPn=311 354453 353575
CT3#1 hypothetical protein
DMRAEMAVIYWDRSKIVWSFEPWSLRLTWYGVFFTVGIFLACLSARYLALSYYGLKDHLS
FSKSQLRVALENFFIYSILFIVPGARLAYVIFYGWSFYLQHPEEIIQIWHGGLSSHGGVL
GFLLWAAIFSWIYKKKISKLTFLFLTDLCGSVFGIAAFFIRLGNFWNQEIVGTPTSLPWG
VVFSDPMGGVQGYPVHFVQLYEGISYLVVSGILYFLSYKRYLHLGKGYVTSIACISVAFI RFFAEYVKSHQGKVLAEDCLLTIGQILSIPLFLFGVALLIICSLKARRHRSHI

354518 354976

CTIO1 hypothetical protein
CTMARNIKYFLILFPGILWISAGMKLLLKATAIALDPLSSFFTYCLLSMVSWGLASLKHR
YLLSKTIRKQLSLSSFFSGKITWIAYIKQTFISRRFLIMVIMIAFSLVLRRYISNPQAL
FVIRATVGYALIKTAIAYFSKLQNALMENPEGN

354957

acpS-Acyl-carrier Protein Synthase
WKILKEISANSMEIIHIGTDIIEISRIREAIATHGNRLLNRIFTEAEQKYCLEKTDPIPS
FAGRFAGKEAVAKALGTGIGSVVAWKDIEVFKVSHGPEVLLPSHVYAKIGISKVILSISH CKEYATATAIALA

356285

CPn_0314 356285 355353 trxB-Thioredoxin Reductase MIHSRLIIIGSGPSGYTAAIYASRALLHPLLFEGFF3GISGGQLMTTTEVENFPGFPEGI LGPKLMNNMKEDAVRFGTKTLAGOLISVDFSVRPFILKSKEETYSCDACIATGSAKRL EIFOAGNDEFWOKGVTRGAVCDGASPIFKNKDLYVIGGGDSALEEALVLTRYGGHVYVVH RRDKLKASKAMEARAONNEKITFLEMSEIVKISGDIVENSVDIKNKYTGEITTREAAGVF FAIGHKPNTDFLIGGLTLDESGYIVTEKGTSKTSVPGVFAAGDVQDKYYRQAVTSAGSGC

358716 ral at Ribosomal Protein

MPROAFYTM SIKELLON LECUTEDVAEFKOLLYTAHP LTGGEEGODE LOPGA LLKGTVV O UNKDPVVVDVGEKSEGV LEMSET LDGGGGGVALAGEVEVVLOQAT DEEJKVVLGREKATE ORQMEY LLANGERE BLVKGOTTRKVKGGLIVD I MEAFLES GO IDNKS KINLDDYGKVG EFKILK INVERRNI VVGRRELLEART ISKABLIEG I GTGGYRKGVKNI TDPGVPLDLD G EDGETTI LIDMLMKB LBHD. JEWAETNÖEFEA L FF JADKERGBAYYTYF VEHNIMED LEK

KYPPGKRVLGKIVKLLPYGAFIEIEEGIEGLIHISEMSWVKNIVDPSEVVNKGDEVEAIV LSIQKDECKISLGLKOTERNPWDNIEEKYPIGLHVNAEIKNLTNYGAFVELEPGIEGLIH ISDMSWIKKVSHPSELFKKGNSVEAVILGVDKESKKITLGVKQLGSNPWNEIEAMFPAGT VISGVVTKITAFGAFVELQNGIEGLIHVSELSDKPFAKIEDIISIGENVSAKVIKLDPDH KKVSLSVKEYLADNAYDODSRTELDFKDSOGPKERKKKGK

358784 360121

LYPIEIQKIAILEDDKVIAIVVNDADYATVIGKRGINARLISHILDYELEVQRMSEYNKL LEIQRLQLAEFDSPHLDQPLEMEGISKLVIQNLEHAGYDTIRRVLLASANDLASVPGISL ELAYKILEQVSKYGESKVDEKPEIED

CPn_0317 360045 infB-Initiation Factor+2 362750 360045

INTÉ-INITIATION FACTOR-2
SLLIRSLSKSANMEKVKLTKNIKLKIKNAQLTKAAGLDKLKQKLAQAGSSEAKSSSEKPS
AKEKSVKVALAATSTPTASAEQASPESTSRRIRAKNRSSFSSSEEESSAHTPVDTSEPAP
VSIADPEPELEVVDEVCDESPEVHPVAEVILPEQPVLPETPPQEKELEFKPVKPAEPKSVV
MIKSKFGPTGKHINHLLAKTFKAPAKEEKVVAGSKSTKPVASDKTGKPGTSEGGEONNRE
KQFNPANRSPASGPKRDAGKKNLTDFFDRPSKKSDESLKAFTGRDRYGLNEGGEEDRWRKK
RVYKPKKHYDEASIQRPTHIKISLPITVKDLAAEMKLKASEVIQKLFIHGMTYVVNDILD
SETAVOFIGLEFGCTIDIDYSEGDELCLSNDTVRDEIQSTDPSKLVIRSPIVAFMHVDH
GKTTLIDSLRKSNVAATEAGAITQHMGAFCCSTPVGDITILDTPGHEAFSAMRARGAEVC
DIVVLVVAGDBGIKECTLEAIEHAKAADIAIVVAINKCDKPNFNSETIYRCLSEINLLPE
AWGSSTVITVNTSAKTGEGISELLEMALAOBAVLELKADPSARARGLYESELHKGLGPVA AWGGSTVTVNTSAKTGEGLSELLEMLALQAEVLELKADPSARARGLVIESELHKGLGPVA TVLIQNGSLKLGEALVFNDCYGKVKTMHNEHNELMKEAGPSIPVLITGLSDIPKAGDPFF VVKNEKTARDIIEARSAGQORFALQOKKRPNFDSMLQNKKTLKLMIKADVQGSIEALVSS ISKIKSEKVDVEILTNSVGEISESDIRLAAASKAVLIGFHTGIESHAEPLIKSLGVRVEL FTVIYHAIDAIKEIMTSLLDPIAEEKDEGSAEIKEIFRSSQVGSIYGCIVTEGIMTRNHK VRVLRNKEILWKGTLSSLKRVKEDVKEVRKGLECGILLEGYQQAQIGDVLQCYEVIYHPQ

CPn_0318 362704 363126
rbfA-Ribosome Binding Factor A
WMSYNVMKLSIIHKNYNLKYCMTENRRIKRVNALLQEAIAKVILKDVKHPKISNLWITVT
RVSLSKDLHSARVYVSVMPHENTIKEEALEALKVSAGFIAHRASKNVVLKYFPELHFYLDD **IFSPODYIENLLWOIGEKEKS**

CPn_0319 363133 363879
truB-tRNA Pseudouridine Synthase
TIFFGNLNTIKDMTMDLAVELKEGILLVDKPQGRTSFSLIRALTKLIGVKKIGHAGTLDP FATGVMVMLIGRKFTRLSDILLFEDKEYEAIAHLGTTTDSYDCDCKVVGRSKKIPSLEEV LSAAEYFQGEIQQLPPMFSAKKVQGKKLYEYARKGLSIERHHSTVQVHLQITKYEYPLLH FVVSCSKGTYIRSIAHELGTMLGCGAYLEQLRRLRSGRFSIDECIDGNLLDHPDFDISPY LRDAHGNSL

363824 364783

ribF-FAD Synthase

TTPISIFLPTYEMPMEIAYSLTSSFSVDSVTVGFFDGCHLGHSNLLSILTSYSGSSGVIT TDSHFQTVLSLNHTKLINTKEERLQLLQTFPIDWLGVLTFDLNFANQSAEEFLTLLHRNL KCKKLILLGYDSCIGKEQQSNFEALDTIGK PLG IEVIKIPPYRMDNIVVSSKAIRQFLSAG NLECAHRFLGHPYAISGKITEGSGIGGSLGFATINLPREESLIPLGVYACEIRYDSTTCQ GVMNLGTAPTFGRESLYAEAHIFSFAENLYGKEVSIIPRKFLREEKKFQSKETLIRAIEK DILDAQDWFAKGSFNYEGTA

365900 364767

yskkhviififrclmshtecgivglpnvgksglfnaltgaqvascnypfctidpnvgivp viderlealakisnsqkiiyadmkfvdiaglvkgasdgaglgnrflshirethaiahvvr CFDDPDVTHVSGKVNPVEDIEVINLELIFSDFSSAKNIHSKLEKLAKGKREVGALLPLFD TIIAHLEKGLPLRTLELTPEQIVALKPYPFLTMKPMFYIANVDESSLPDMDNDYVAAVRE VAAKENSKVVPICVRIEEEIVSLPIEERLEFLMSLGLEKSGLHRLVRAAYDTLGLISYFT TGPQESRAWTVVRGSSAWEAAGEIHTDIQKGFIRAEVITFEDMIECQGRAAARELGKLHI EGRDYIVODGDTMLFLHN

366231 367328

CPH_0322 366231 367328
yscU-YopS Translocation Protein U
SNLONSMGEKTEKATPKRLRDARKKGOVAKSODFPSAVTFIVSMPTAFSLSTFFFKHLGG
FLVSMLSQAPTRHDPVITLEFYLKNCLMLILTASLPLLGAVAVVGVIVGPLIVGPTFSTEV
FKPDIKKFNPIENIKQKFKIKTLIELIKSILKIFGAALILYITLKSKVSLIIETAGVSPI
ITAQIFKEIFYKAVTSIGIFFLIVAILDLVYGORHNFAKELKMEKFEVKQEFKDTEGMPEI
KGRRRQIAQEIAYEDSSSQVKHASTVVSNPKDIAVAIGYMPEKYKAPMIIAMGINLRAKR
ILDEAEKYGIPIMRNVPLAHQLLDEGKELKFIPESTYEAIGEILLYITSLNAQNPNNKNT

CPn 0323 367322 369460

CPn_0323 367322 369460
lcrD- Low Calcium Response D
SFIMMKLIMFVSRTLGGDTALMMINKSSDLILALWMMGVVLMITIPLPPPIVDLMITINL
SISVFLLMVALYIPSALQLSVFPSLLLITTMFRLGINISSSRQILLKAYAGHVIQAFGDF
VVCGNYVVGFITFLITTIGFIVVTKGAERVAEVAARFRLDAMFOKOMAIDADLRAGMID
ATOARDKRAQIOKESELVGAMDGAMKFIKGDVIAGIVISLINIVGGLTIGVAMHGMLAQ
AHVYTLLSIGDGLVSOIPSLLIALTAGIVTTRVSSDKNTNLGKEISTQLVKEPRALLLA
GAATLGVGFFKGFPLWSFSILALIFVALGILLETKKSAAGKKGGGSGASTTVGAAGDGAA
TVGDNPDDYSLTLPVILELGKDLSKLIGHKTKSGOSFVDMIPKMRQALYODIGIRYPGI
HVRTDCPSLEGYDYMILLNEVPYVRGKIPPHHVLTNEVEDNLSRYNLPFITYKNAAGLPS
AWVCEDAKAILEKAAIKYWTPLEVIILHLGYFFHKSGQEFUGIQEVRSMIEFMERSFPDL
VKEVTRLIPLOKLTEIFKRLVGEDISIKDLEFTILESLGEWAQTEKDTVVLLTEYVRSSLKL
YLGFFFSGOQGAISVVLLDPEIEEMIRGAIKYTGASSYLALDPDGVNLILKSMRNTITFT
PAKKOPPVLLTAIDVRRYVRKLIETEFPDIAVICYQEILPEIRIQPLGRIQIF

369468 370688

CPG_0324 SP0468 SP0468

AEGIGTENDUNFVKVAEGYHKIINDKEPTAGKVEREVRNLIGDDVDGVTGVUNLFFGALR QTGSRUFGGADKRQQUGAMIANALDAVNINNEDYPKASDFPKPYPWS

170696 371148

CPN_U325 370696 371148 CT325 hypothetical protein KRIAMONQYEQLLESLAPLLNTTLAPDKNNSCLIRFSDTHVPVQISEDGNSGDLAVSTLL GTLPENVFRERIFKAALSVNGSFQSSIKGILGYGEVTQQLYLSDILSMNYLNGEKLFEYL KLFSLHAKIWMESLRTONLPDLHVLGIYYVA

m 0 . 37118

malQ-Glucanotransterase

malQ-Glucanotransterase
PSCFGNLLERVNULKYTKHSPSAHAWKLIGTSPKHGIYLPLFSIHTKNSCGIGEFLDLIP
LISWCQKQGFSVIQLLPLNDTGEDTSPYNSISSVALNPLFLSLSSLPNIDTIPEVAKKLQ
DMHELCSTPSVSYTQVKEKKWAFLREYYQKCCKSSLEGNSNFSPELESERYWLYPYCTFR
AIKHHMHGPINNWPKSLTDQENFPDLTKKFHDEVLFFSYLQFLCYQQLCEVKAYAQQHH
VLLKGDLPILISKDSCDVWFRDYFSSSRSVGAPPDLYNSEGQNWHLPYNFSQLAKDDY
VLLKKEDLPILISKDSCDVWFRDYFSSSRSVGAPPDLYNSEGQNWHLPYYNTEISTWIG IMWKERLEYAQNEYSYYRLDHIIGFFRLWIWDSSGRORFIPDNPKDYIKQGTEILSTMLG ASSMLPIGEDLGIIPQDVKTTLTHLGICGTRIPRWERNWESDSAFIPLKDYNPLSVTTLS THDSDTFAQWWLNSPKBAKOPAKFLHLPPOKTLTTETQIDILKLSHESASIFHLTMLFNDY LALCPDLVSKNLQRERINTPGTISKKNWSYRVRPSLEELAIHKKFNGYIEKILTGL

372927 373211

CPH_032/ **128-L28 Ribosomal Protein RIHKKNMSRKCPLTGKRPRRGYSYTLRGIAKKKKGIGLKVTGKTKRRFFPNMLTKRLWST EENRFLKLKISASALRHIDKLGLEKVLERAKSKNF

CPn_0328 373220 374992
CT085 hypothetical protein
LKYREIFMSFLRRHISLFRSCKQLIDVFAPVSPNLELAEIHRRVIEDQSPALLFHNVIGS
SFPULTNLFGTKHRVDQLFSQAPODNLIARVAHLISSTPKLSSLWKSRDLLKRISSLGLKK
ARFRRFPFVSMSSVNLDHLPLLTSWPEDGGAFLTLPLVYTESPTLTTPNLGMYRVQRFNQ
NTMGLHFQIQKGGGMHLYEABQKKQNLPVSVFLSGNSFLUTLSAIAPLPENVSELLFATFL
QGAKLLYKKTNDHPPLLYDAEFILVGESPAGRRPFEDFFGDHFGYYSLQHDFPEFHCHK
IYHRKDAIYPATVVGKEYQEDFYIGNKLQEYLSPLFPLVMPGVRRLKSYGESGFHALTAA VYKERYMESLITALRILGEGOLSLIKFLMYTOEVPLDRFSVVLETILERLOPDRDLII FSETANDTLDYTGPSLNKGSKGIFMGIGKAIRDLPHGYQGKIHGYQDIAPFCRGCLVLE TSLEDRCIKSLLHHPDLKSMPLIILADNLRETIOSEKDFLWRFFTRCAPANDLHALHSHF ATHRPNYNFPFVIDALMKPSYPKEVEVDPSTKQKVSERWHAYFPNKETFYI

375085 376146

CPA_0329 375085 376146
Phopholipase D Superfamily [leader (33) peptide)
KMRKRQKDKLKICVIISTLILUGIFARAPRGDTFKTFLKSEEAIIYSNQCNEDMRKILCD
AIEHADEEIFLRIYNLSEPKIQQSLTRQAQAKNKVTIYYQKFKIPQILKQASNVTLVEQP
PAGRKLMHQKALSIDKKDAWLGSANYTNLSLRLDNNLILCMHSSELCDLIITNTSGDFSI
KBÖĞTGKYFVIPQDRKIAIQAVLEKIQTAQKTIQVAMFALTHSEIIQALMQAKORGHYDI
IİDRSHSKLTFKQLRQLNINKDFVSINTAPCTLHHKFAVIDNKTLLAGSINWSKGRFSLN
DESLIILENLTKQQNQKLRMIWKDLAKHSEHPTVDDEEKEIIEKSLPVEEQEAA

Ein 330 376930 376202
CTO 330 376930 376202
CTO 33 hypothetical protein
FISIEMLLLSROLFSVLPSRFQDLHVYRFKESLKLLQFMTMVGGEIVVVLAEIKEEDLRA
RKEËVRKREKNYLRIFRVLSRFDVMRIIRFDPYGALSAQSIAKDSRQNSPLVEKISEEI
ATNEAIRLAILAIGDREQEEKKQRHRYKLLGOKQAKVLLSQLRHVHLDFKKLYCDSKKKE
DØEKDEKNKQKRSIKVTKKKKKGISLGAAASQAIAAAAEAWVIARNKGVLETASTLFYQKD
TPA.

CEni0331 378452 376701 CT082 hypothetical protein IGRIMAVSGGGGVQFSSDFCKWNPALQGEQAEGPSPLKESIFSETKQASSAAKQESLVR SGSTGMYATESQINKAKYRKAQDRSSTSPKSKLKGTFSKMRASVQGFMSGFGSRASRVSA KRASDSGEGTSLLPTEMDVALKKGNRISPEMQGFFLDASGMGGSSSDISQLSLEALKSSA KRABDSGEGTSLLPTEMDVALKKGNRISPEMQOFFLDASGMSGSSSDISQLSLEALKSSA
FSGARSLSLSSSESSSVASFGSFQKAIEPMSEEKVNAWTVARLØGEMVSSLLDPNVETSS
LVREMATGNEGMIDLSDLQQEEVSTAMTSPRAVEGKVKVSSSDSPEAMPTGIPNSNTLE
PAEKEAEKQESREQLSEDQMMLARAMAGLLTGAAPQEVLSNSVWSGPSTVFPPPKFSGTL
PTGESCDKSKHKSPGIEKSTNHTHFSPLREGTVKSAEVKSLPHPESMYRFPKDSIVSREE
PEAVVKESTAFKNPENSQNFLPIAVESVFPKESGTGGALGSDAVSSYHFLAQRGYSLL
APLPRATDDYKEKLEAHKGPGGPPDPLIYQYRNVAVEPPIVLRSPQPFSGSSRLSVQGKP
EAASVHDDGGGGNSGGFSGDQRRGSSGQKASRQEKKGKKLSTDI

CPn_0332 378676 378536
CHLTR T2 Protein
YLDSRIRVIPLARQRCTLLHLLAVLCPPISFFTQGVSPCVFFCFLDF

378800 CPn 0333 379117

TUUB VDFFVFVFFMGKPKKSRTDRALAQEIQKKSTEVLKKPARIKAKNRRKFLIAKEQKTLKHR AQEYDQLVRSLLDSQKKDTDKVLIFNYENGFVFTDKDHFSKYSIRL

CPn_0334 379306 379823 CTO79 similarity TMSVHITPRKCFILCILSMFTLPTLFPKAHLILFSPYIVLCFYCFSKDKGLVLALGCGVL SDLALGSRGVFLLLYPLTALITHRAHLIFSKESKAALVIVNMIFYGVFLLLTIPMCALFG HEVRWSIDVLMIPLKCSFLDNLIFTSVIYILPCAINSGIHKMISFFRRLVCY

379808 380674

TOTO-METHYLENE TETTAHYDTOFOLISE DEHYDTOGENASE
EIGMLLRGIPAAEKILORLKEEISGSPTSFGLAVVLIGNDPASEVVVGMKVKKATEIGII
SKAHKLPOSTLSSVIKLIERLNQDPSIHGILVQLPLPKHLDSEVILQAISPDKDVDGLH
PVNMGKLLLGNFDGLLPCTPAGIIELLNYYEIPLRGPHAAIVGRSNIVCKPLAALMMGKH POTNCTVTVLH::QSENLPETLKTADTTAALGAPLFTKETMVAPHAVTVDVGTTRVPADN AKGYTLL::DVDFNNVVTKCAATTPVPGGVGPMTVAMLMSNTWRCYQNFS

380569 381591

YOLD
REDMMPCHERODWRRW. HEXTRY ALLOYMAMLER PELVELOLOLOGO SCOOKTTT LEGEOMT L
FYR EVEZTSE: AKEKARLEGO LOROFHK LOG LYNNMINEYSELG LINRAPADVE LTEDVEL
DEFELOGVERLYKEDEORFET PVGELKTEURLEHEKSOTLE PROTVEROHYKEMGROHLE POS
FIRKTELLYKEN HIVO LOELANVYKKYYAND CLUBELGENFOLENNYVERKEELKT JOHHESGREWR
LEDFAACT LED LOMA LATDOMHLOKWEVEGK LYTH LEDTRYGKELEELSSY PLOSYSVYH
ECCAYADA LATVEMTPONK LEAKOWAREHH LETY LIDGACS

CPn_0337 392141 381575
smpB- Small Protein B
IEEIFPGNQGKRILIIVLRPKNCFLLYWTLSPIMGEDLMAQKEIVSNRKALRNYEVIETL
EAGIVLTCTEIKSLRDHGGNLGDAYVIVSKGEGWLLNASIAPYRFGNIYNHEERRKRKLL
LHRYELRKLEGKIAQKGMTLIPLGMFLSRGYVKVRLGCCRGKKAYDKRRTIIEREKEREV

CPM_0338 382272 383375

CPATATA IN DOTAGE TO THE TOTAL OF THE CONTROL OF THE TOTAL OF THE CONTROL OF THE CONTRO MPMRLHDD

CPn_0339 383405 384034
CT339 hypothetical protein
VTTLPMFMKICSLKLKNFRHSDLEISLAPKLNYAQGKTNLLEALYVLSLGRSFRTQHLT
DTIFFGSSHFFLETOFEKDHLPQALSIYTDKQGKKICYNQLPIKTLSQLIGKVPIVLFSS
KDRLLISGAPADRRLFLINLLLSQCDNHYTLCLSYYHRALQQRNALLKSKQTSTVASGMNS WSNTAPTYPSNGFSVVRNFQIYPKNFGLTT

(frame-shift with 0339)
PLYPLLIVLSSRSSAEKCSLKKQANLNRGLWDEQLVKHGTYLSIQRFLCSQKLSDLSKEL WSNNLKEQLALKFKSSLIKNSDISETAVAEEFHKQLSISLPRDLE

384160 384495 CPn_0341

(frame-shift with 0340)
GSTSVGPHREDFLLTMNQMPVSQFSSEQQKHSLLAILRLAECLYLKQSHHVSPLVCLDDI HAGLDNERVGQLLDPAPTLGQTLITSTHMHGELPKTSLVLSIENAQVSEQII

384619 385062

CPT_U342 384619 355062
predicted OMP [leader (19) peptide]
HMKKFLLTILFLAVGNPLFSETSVIOTLPSGIGGLKETSKOKESVVCVHAFLRSYTSLKP
LARVLEKEHYDVFIWNYETRKFTLEKHAEHLNRLLKKIAELKPGVPINFVTHSIGGVIVR
VALAHPDCPEEAKKEKLFSWLLRTQGLH

384999 385595 CPn_0343

(Frame-shift with 0342?)

LPRSQKRKAILMAPPNAGSTLARRYRCVKFVQFVFGGKLGRQLLTYCPTKMLNVGKLPS

SLDVLILGSQKRKSKLEPFRLPYENDGKVCTIETKLDTPHKAYVIHTSHTYIITNKSLYL

MKEFLKEGNTTPIIEHVPEAALEQTVMEDKQKNSRLKPYPNQDIYVIHCFGSRPYNLYGF PKKWSLNOKNE INPEKLEK

387432 385558

yaeL-Metalloprotease
SSRYMTITYFILAALALGILVLIHELGHLVVAKAVGMAVESFSIGFGPALFKKRIGGIEY SSRYMTITYFILAALALGILVLIHELGHLVVAKAVCMAVESFSIGFGPALFKKRIGGIEY
RIGCIPFGGYVRIRGMERTKEKGEKGKIDSVYDIPQGFFSKSPWKRILVLVAGPLANILL
AVLAFSILYMNGGRSKNYSDCSKVVGWVHPVLQAEGLLPGDEILTCNSKPYVGDKDMLTT
SLLEGHLNLEIKRPGYLTVPSKEFAIDVEFDPTKFGVPCSGASYLLYSNQVPLTKNSPME
NSELRPNDRFVWMDGTLLFSMAQISQILNESYAFVKVARNDKIFFSRQPRVLASVLHYTP
YLRNELIDTQYEAGLKGKWSSLYTLPYVINSYGYIEGELTAIDPESPLPQPGERLQLGDR
ILAIDGTPVSGSYDILRLVQNHRVSIIVQQMSPQELEEVNSRDADKRFIASYHSEDLLQI
LNHLGESHPVEVAGPYRLDPVVGPPMIDVYSSESLDKQLEVAKKIKNKDKORYYLERLD
AEKQKPSLGISLKDLKVRYNPSPVVMLSNITKESLITLKALVTGHLSPQWLSGPVGIVQV
LHTGMSVGFSEVLFWIGLISMNLAVLNLLPIPVLDGGYILLCLWEIVKRRRLNMKIVERI
LVPFTFELLIIFFIFLTFODLFFFFG LVPFTFLLIIFFIFLTFQDLFRFFG

388587 387436

CPT_0345 sypothetical protein
LKVACLKHLAVLGSTGSIGRQTLEIVRRYPSEFKIISMASYGNNLRLFFQQLEEFAPLAA
AVYNEEVVNEACQRFPHMOFFLGGEGLTQLCIMDTVTTVVAASSGIEALPAILESMKKGK
ALALANKEILVCAGELVSKTAKENGIKVLPIDSEHNALYQCLEGRTIEGIKKLILTASGG
PLINKSLEELSCVTKQDVLNHPIWNMGSKVTVDSSTLVNKGLEIIEAYWLFGLENVEILA
VIHPOSLIHGMVEFLGGSVISIMNPPDMLFPIQYALTAPERFASPRDGMDFSKKQTLEFF PVDEERFPSIRLAQQVLEKQGSSGSFFNAANEVLVRRFLCEEISWCDILRKLTTLMECHK VYACHSLEDILEVDGEARALAQEI

CPn_0346 389690 388704 070-troD/ytgD-integral Membrane Protein KKGSLMALGPSPYYGVSFFQFFSVFFSRLFSGSLFTGSLYIDDIQIIVFLAISCSGAFAG TFLVLRKMAMYANAVSHTVLFGLVCVCLFTHQLTTLSLGTLTLAAMATAMLTGFLIYFIR NTFKVSEESSTALVFSLLFSLSLVLLVFMTKNAHIGTELVLGNADSLTKEDIFPVTIVIL ANAVITIFAFRSLVCSSFDSVFASSLGIPIRLVDYLIIFOLSACLVGAFKAVGVLMALAF LIIPSLIAKVIAKSIRSLMAWSLVFSIGTAFLAPASSRAILSAYDLGLSTSGISVVFLTM MY IVVKFISYFRGYFSKNFEK I SEKSSQY

CPn_0347 391078 389678

069-troC/ytgC-Integral Membrane Protein
TFGTNPEALSRRTIWIVLINLSCVFSDTIFLSSFLAVTLICMTTALWGTILLISKQPLLS
ESLSHASYPGLLVGALMACYVFSLQASIFWIVLFGCAASVFGYGIIVFLGKVCKLHKDSA
LCFVLVVFFAIGVILASYVKESSPTLYNRINAYLYGQAATLGFLEATLAAIVFCASLFAL
www.rQI'VVTTFDKDFAVTCGLKTVLYEALSLIFISLVIVSGVRSVGIVLISAMFVAPSL
GARQLSDRLSTILLISAFFCGISGALGSYISVAFTCRAIIGQAVPVTLPTGPLVVICAG
LLAGLCLLFSPKSGWVIRFVRRYHFSFSKDDEHLLKVFWHISHNRLENISVRDFVCSYKY
QEYFGPKPPPRWRVQILEWRGYVKKEQDYYRLTKKGRSEALRLVRAHRLWESYLVNSLDF
SVEGVHELAEEIEHVLTEELDHTLTEILNDPCYDPHPQIIPNKKKEV

CPn_0348 301815 391027
068-troB/ytgB-ABC transporrer ATPase
FOWLLNYPDETFWSVRNLCVNYEHAAVLYHTDFDLDYGGUTAILOPNGAGKOTLLKASLG
LIPPSGTTYFFNORKEKUROP TAYMPORADUMDFPMTVLDLALMGRYGYKGMWGRISS
DDFREAFHILERVGLEGVAPRO COQLOGOQOOPAFLARALMQKADLYLMDELFGAIDMAS
FYTTSVGVLQELROQGKTIVVYHHDLCHVRQLFDHVVLLNKRLICGGPTDEGLNGDTIFQT
VGCTIFLL FOTIKLEGKNDGG YOCETELLEOTLKLSRGKOFGSC

(Pn_0349 3NCA) 191790 967-troA/ytgA-Solute Protein Binding Family

WILKNA:REMDAKMGYIFKVMRWIFCFVACGITFGCTNSGFQNANSRPCILSMNRMIHDC VERVVGNRLATAVLIKGSLDPHAYEMVKGDKDKIAGSAVIFCNGLGLEHTLSLRKHLENN PNSVKLGERLIARGAFVPLEEDGICDPHIWMDLSIWKEAVIEITEVLIEKFPEWSAEFKA NSEELVCEWSILDSWAKQCLSTIPENLRYLVSGHNAFSYFTRRYLATPEEVASGAWRSRC LSPEGLSPEAQISVRDIMAVVDYINEHDVSVVFPEDTLNQDALKKIVSSLKKSHLVRLAQ KPLYSDNVDDNYFSTFKHNVCLITEELGGVALECQR

393169 393684 CPn 0350

THE TOTAL STATES OF THE STATES

393861 395432

CPn_0351 393861 395432
adt-ADP/ATP Translocase
kikvporvmmtkteekpfgklrsflwpihthelkkvlpmflmffcitfnytvlrdtkdtl
lvgapgsgabaipfikfwlvvpcatiffmliyaklsnilskqalfyfavdtpffliffalfpt
viyplrdvlhptefadrlqailppgllglvailrnwtfaafyvlablwgswmlslmfwgf
aneitkiheakrfyalfgiganisllasgraivwasklrasvsedvdfwgislrllmamt
lvsglvlmasywwinknvltdprfynpeemokokkgakpkmmmkdsflylarspyillla
tlviavgicinlievtwksqlklqypnmndyseffmgnfsfwtgwsvlimlfvggnvirk
fgwltgalvtpvmvlltgivffalviffnqasglvamfgttplmlavvvgatqnilskst
kyalfdstkemayipldqeokvkgkaaidvvaarfoksggaliqqgllvicgsigamtpy
lavillfiiaiwLvsatklnklflaqsalkeqevaqedsapass

CPn_0352 395478 396830 No robust homolog present in Genebank/EMBL as of 11/7/98 WVGIFFINSHFTNSYAFFNQKVIITVRHSGCTMKCSPLTLVPHIFLKNDCECHRSCSLKI WVGIFTINSHI INSTAFTNONTITY WITHOUT HEAVING THE LIVERT ENDEL HIS WITH ENDEL HIS WITH END ALL SELECTION OF THE WITH END ALL SELE QFLFLFFSHGITWEQAQMIQLINPDNWKMLCQFDKAGGHCSMATFGGFLNTETNMFDPVS SNYEPTVNFMTWKELKVLLEKVKESPMHPASALVQKICVNTTHHQNLLKRWQFVRNTSSQ WTSSLPQYAFHAQTYKLEKKIESSLPIRSSL

CPn_0353 396893 397135
No robust homolog present in Genebank/EMBL as of 11/7/98
LRFRNIKKSLIFIKRIRYSQSGKEQKGARPFFKKSITSSLVILLLEAIFNENFSSIIQNN FNKNFKNKNISINRIFVKFTI

CFT_0354 397062 398507

NO_FODUST homolog present in Genebank/EMBL as of 11/7/98
YKY#SIKILKIKTFLLIGFLINLRYNTQIDEPRKCMSNITSFVIQNIRSCNYYFELKNST
THHTVISAILLCGAL IAFLCVAAPUSYLLSGALLGLGLILALIGVILGIKKITPMISSKE
QVFFQELVNRIRAHYPKFVSDFVSEAKPNLKDLISFIDLINQLHSEVGSSTNYNVSEELQ
QKILDFFEGIARLKNEVRTASLKRLESAASSRPLFPSLPKILQKVFPFFWLGFISAGSKV
VELHEVKHIGGSLEEDLSDYIKPEMLPTWHI PLDFRPFTNSSILMLTUVLARVLTRDVF
QHLKYAALNGEWNLNHSDLNTMKQQLFAKYHAAYQSYKHLSQPSLQEDEFYNLLLCIFKH
RYBWKQMSLIKTVPADLWENLCCLTLDHTGRPQDMEFASLIGTLYTQGLIHKESEAFLSS
LTHESLDOFKTIROSTNIAMFLENLATHNSTFSLPPITYHPLKRYVFSOPEEDESSLL LTEESLDQFKTIRRQSTNIAMFLENLATHNSTFRSLPPITVHPLKRSVFSQPEEDESSLL

CPhe0355 39995 398591

No Tobust homolog present in Genebank/EMBL as of 11/7/98

IRDFYLHIIYTAFNRSISKELAMSMTIVPHALFKNHCECHSTFPLSSRTIVRIAIASLFC
IGALAALGCLAPPVSYIVGSVLAFIAFVILSLVILALIFGEKKLPPTPRIIPDRFTHVID

EAYGLSISAFVREQQUTLAEFRQFSTALLCNISPEKKIKQLPSELRSKVESFGISRLAGD
LEKNWPIFEDLLSQNCPLYWLQKFISAGDPQVCRDLGVPRECYGYYMLGPLGYSTAKAT
IFÈKETHHILQQLTKEDVLLLKNKALQEKWDTDEVKAIVRECYGYYMLGPLGYSTAKAT
KEŤISKELLLLSLHGYSFDQLQLITQLPRDAWDWLCFVDNSTAYNLQLCALVGALSSQNL
LDĒŠIDPDVNLGLYVIQDLKEAVQAFSASDEPKKELGKFLLRHLSSVSKRLESVLRQGL
HRIALSHGNARARVYDVNFVTGARIHRKTSIFFKD HRÏALEHGNARARVYDVNFVTGARIHRKTSIFFKD

CPn_eq356 400465 400109
No Hebust homolog present in Genebank/EMBL as of 11/7/98
KQVQLFQYMNESGWDWLCDFDSQGEGFQLSRLVGLLHSSWALYEAKEQFYLPEVSLLTWE
ELFEMQLLSKPTKHGVAKDLCNVFEKHFQRFRQYLGSLDLNQRFENTFLNYPKYHLDRE

CPn_0357 401341 400469
No robust homolog present in Genebank/EMBL as of 11/7/98
YSSHNGASMVNIQPVYRNTQVNYSQATQFSVCQPALSLIIVSVVAAVLAIVALVCSQSLL
SIELGTALVLVSLILFASAMFMIYKMRQEPKELLIPKKIMELIQEHYPSIVVDFIRDQEV
SIYEIHHLISILNKTNVFDKAPVYLQEKLLOFGIEKFKDVHPSKLDPFEELLLOHCPLHW
LGRLVYPMVSDVTPGTYGYYWCGPLGLYENAPSLFERRSLLLLKKISFGEFALLEDGLKK
NTWSSSELVQIRQNLFTRYYADKEEVDEAELNADYEQFDSLLHLIFSHKLS

401757 401578

CPR_U378
No robust homolog present in Genebank/EMBL as of 11/7/98
EEVLSVSMKLIPTQDSIERETDSKRDKKIFTIYICSSKVLAGHFFSHLDKHNKIHESIGV

CPn 0353 401994 403817

CPn_0359

LepA-GTPase

ITLQY!LKEYK!ENIRNFS!IAHIDHGKSTIADRLLESTSTVEEREMREQLLDSMDLERE
RGITIKAHPVTMTYLYEGEVYQLNLIDTPGHVDFSYEVSRSLSACEGALLIVDAAQGVQA
QSLANVYLALERDLE!IPVLNK!DLPAADPVRIAQQ!EDY!GLDTTN!IACSAKTGQIP
AILKA!IDLVPPPKAPAETEKALVPDSHYDPYVG!MVYVR!ISGELKKGDRITFMAAKG
SSFEVLG!GAFLPKATF!EGSLRFQQVGFF!ANLKKVXDVK!GDTVTKTKHPAKTPLEGF
KE!NPVPFAG!YP!DSSDFDTLKDALGRUQLNDSALT!EQESHSLGFGFRCGFLGLLHL
E!!FFP! I! REFDLD!!ATAPSO!YKVVLKNGKVLD!DNPSQYPDPA!!EHVEEPWVHVN!
ITPQEYLCN!MNLCLDKRG!CVKTEMLDQHFLVLAYELPLME!VSDFNDKLKSVTKGYGS
FDYRGJDYPKGS!KLEVL!NEEP!DAFSCLVHRDKAESRGRS!CEKLVDV!PQUFK!P
IOAATINKV!ARET!RALSKNVTAKCYGGD!TRKRKLWEKQKKGKKRMKEFGKVS!PNTA
P!EVLKLD FIEVERED

CFm_0340 405361 40.032 CF360 hypothetical protein VALQMHGLIGLAVMCKNLVINMIDHGESVSVVNPTPEKTRDFLKEYPNHRELVGFESLE DEVELORERERKIMEMIQACKEVDOSTHALLPTLERGDVITDGGGSVFRDGERECKELQEK GLUPIZZGLGGGERGARHGPGIMEGGNPEAWPLVAPIFQSIAAKVQGRECGWYGTCGAG HYVKAVHNOTEYODTOLICEAYOTTRDI'LKLSATAVATTLKEWNTLELESYLTRTASEVL

ALKDPEGIPVIDTILDVVOGKGTGKWTAIDALNSGVPLJULIGAVLARFLIJWKEIREQA ARNYPGTPLIFEMPHDPSVFIQDVFHALYASKIIGYAQGFMLLDEASKEYNWGLDLGEIA LMWRGGCIIQSAFLDVIHKGFAANPENTSLIFQEYFRGALRHAEMGWRRTT/VTAIGAGLP IPCLAAAITFYDG/PTASSSMSLAQGLRDYFGAHTYERNDRPRGEFYHTDWVHTKTTERV

405550 405382 CPn 0361

CPR_USEL

EYES-tyrosyl tena Synthetase

THE NORTH SET OF STREET TO THE FOR A CONTROL OF THE F FYHLFKNYGTILQCGSDOWGNITSGIDFIRKGLGQAYGLTYPLLTNAQGKKIGKTESG TVWLDSDLTSPFELYQYLLRLPDDTIFKIARTUTLLSNEEIQDIDRRVQTDFVAVKEFVA QDILSAIHGDLGLEEALSVTRSMHPGNLSSLSEKDFHELFAGGMGASLDKSEVLGKRWLD LFLVLGLCKSKGEIRRLIEQKGVYINNVPIANEHSVCEEQDICYGHYVLLAQGKKRKLVL

CPn_0362 407843 407055
fliA/rpsD-Sigma-28/WhiG Family
LDKKKFVKTQQTQNIIEV-WPYWETQEIEYRDSLIEFYLPLVKSVVHRLISGMPSHVKTE
DLYASGVEGLVRAVERYNPERSRRFEGYAVFLIKAAIIDDLRKQDWVPRSVHQKANKLS
AMDSLRGSLGKEPTDLELCEYLHISQOELSGMFVSARPALIVSLNEEWPSQSDEGAGMAL
EERIPDERAETGYDVVDKQEFSLCLANAIQELEEKERKVMALYYYEELVLKEIGKVLGVS
ESRVSQIHSKALLKLRAALSAFR

409700 407943

CPn_0363 409700 407943
flhA-Flagellar Secretion Protein
EAVFYSGKKDGVRGMIFVPLSILVLIFLDPLOILLDFGLCISFALSLLTVCWVFTLNSSN
SAKLFPPFFLYLCLLRLGLNLASTRWIVSSGTASSLIVSLGSFFSLGSLWAATFACLLLF
FVNFLMVSKGSERIAEVRSRFFLEALPAKQMALDSDLVSGRASYKAVKKQKNALIEEGDF
FSAMEGVFRFVKGDAIISCILLLVAVVSVTCLYYTSGYALEGQMFFTVLGDALVSGVPALL
TSCAAATLISKIDKEESLLNYLFEYYKQLRQHFRVVSLLIFSLCCIPSSPKFPIVLLASL
LWLAYRKEEPASEDSCIERAFSYVEGACPKEQESOFYQVYRAASEEVFEDLGVRLEPVLTS
LRIEERPWLRVFGQNVYLDEMTPEAVLPFLRNIAHEALNAEVVQKYLEESERVFGIAVED
IVPKKISLSSLVVLSKLLVRERVSLKLFPKILEAVAVYQNSGDSLEILAEKVRKSLGYWI
GRSLMDQKGTLEVITIDFHVEELINSSYSKSNPVMQENVIRVDSLLERSVFKDFRAIVT
SCETRFEMKKMLDPHFPDLLVLSHDELPKEIPISFLGIVSDEVLVP

CPn_0364 409954 410238 fer4-Ferredoxin IV KENSMAKLVITSDDEQQEFELEDNSEIAEPCESMGIPFACTEGVCGTCVIEVLEGRENLS EFTEPEYDFLGEPEDSNERLACQCRIKGGCVKVTF

CPn_0365 410498 411544

No robust homolog present in Genebank/EMBL as of 11/7/98

FKGTQVNSLIMATISPISLTVDHPLVDTKKKSCSNFDKIQSRILLITAIFAVLVTIGTLL

IGLLINIPVIYELTGISFIAVVLSNFILLYRRATTLIKPRACGKHKEIKPKRVSTNLQYSS

ISIAINGSKENWEHOPEDLQNLPAPSALLTDNPYEIWAKHSLFSLVSLLPGGNPEHLLI

SASENLGKTLLIEETSQNAPISSYVDTTPSPKSLLNEAIQETRVEINTELPAGDSGERLY

WOPDFRGRVFLPGIPTTPEAIYQYYYALYVTYIQTAINTNTQIIQIPLYSLREHLYSREL PPQSRMQQSLAMITAVKYMAELHPEYPLTIACVERSLAQLPQESIEDLS

CPn_0366 411976 412440
No robust homolog present in Genebank/EMBL as of 11/7/98
MGYLPVSATDVLFESPAAPLINSANTQNQKLIELKGKQQAESSPRTITSVILEVLLVIGC
CLIVLSLLAIRPALGFTLETGHPAAIAVLAVSGTILLVAVIILFCFLAAVPFAAKKTYKY
VKTVDDYASWHSHQQTPTLGTIFSGIVYAESQAQL

CPn_0367 413078 413836

OFTI-USB/ 4130/6 4130/6 4130/6 WISBON TOBUST HOMOSOM PROPERTY OF THE PROPERTY PRLKTSGRVMIVNAANSNMQSGGAGTNAALSAATHPTCWNNTRTSGGKINTGKGLSVGEC RSAPWINRDWINK

CPn_0368 413766 414107 No robust homolog present in Genebank/EMBL as of 11/7/98

TLAKDYLWVNAAQHPGSIETGRINDTNPGEAHFLAQLLGPKYEGELKAHPEKLSNVIKKA YLNCFDEALNNQATVVQVPLISSSIYSPGGKLELEPVNQTKPNSSAYKLYHIRT

CPn_0369 414345 415562 CT058 hypothetical protein_2 NimTosNepleySYDASLYRPPAKHSYPIRLPLNRTDRIEKILKIVTLTLALACALGFSIA AGILAMPIFSAVVVITLAIAAVSLYSLLKKPKLYEILPQIEPESEQSSLSPSPQPPEQQD LPLQIDPLPDPESLPEVSLADLTTPPEELTAITVTPGYEALLEONWOLLPSLAAVDPSFT TETPQQPCFIWKLKDSKLIFISTSGDIAVPRIKTQGRVMIVNAANENISREGGGTNKALS LATSLQCWNASRLFRAHSRSGSQLQPGECRSAKWENSDHTSNDHYPGKAHFLAQLLQPEA AKCNNDPKQAFEVSKKAFHNLFQEAEIIGVDVIQLPLIGCNLFAPSRLLNLGKTRAEWIE AIKLALITSLQDFJWEQDNQEEQKIIILTDKDQPPIIPPRFDLTTP

CPn_0370 415755 416912
CT058 hypothetical protein_3
KRIFFKLFVFYLKSFMSTTEPNLTNVNLTMLISSESMPTQLASHKLKGLDLVAFILIIGI
AVSSGTAAIILGIPLLFILTALAVLAFSILLYFLLREPKSPISVTHQPTPIIKDTDLPPV
PPLALTPVPTEAVLEEPPLFSPRTHQTLLQENWDRIPDLQANTDMPFIAADNOTGYAWHL
KNSNLTLISTLGFIEKPRYKTQGIVMIVNAATPNMANNVKGTSLALAKATSVRCWENSKK SPDPLRSKOPLOLGECRSAKWENLINGTTNAGKAGLPOPLGQLLGPKASDYNYNPNDAFTF CRQAYLNCLNEAKRRKTTVVQLPLLGGHFPGGPKDEETTSLRLQWIDGVKLALIDALQTF GSEAENQNQPWVIILTTLARHPLITP

CPn_0371 417141 417503

No robust homolog present in Generank/EMBL as of 11/7/98

KTMPVSSAPLPTSHRPSSCNLGLMERREKALLARHIODKTTKTIKLLAKTLAALLAILUIEVLG

TIAAFFIPGTPPICLIILGBLILTTYLAVLLLAVIKLLAVIKTIKTTREGOTKRKLSSKSI

CPn_0372 4U'651 4IR061 No robust homolog present in Genetank/LMBL as of IU/7/98 NYRACHRNIGHMSSPVVTGTSCACPVEGTYLGEFLERLGGGGGRCIKIAFAA/TTALLLIN TEVOGIVALAMIFYYTSVGAYFT/IGPLELL BILLIAIMLIGMYKITHPSQNTPISN

CPn_0373 418356 420218

GPn_0373 418356 420218

gcpE

MSEIFEIFMTLITPAINSSPRKTHTVRIGNLYIGSDHSIKTQSMTTTLTTDIDSTVEQIY
ALAEHNCDIVWTVQGIKEAQACERIKERLIALGLNIPLVADIHFFPQAAMLVADFADKV
RINPCNYIDKRNMFKOTKIYTPASVAQSLLRLEEKFAPLVEKKRIKJKAMRIGSVHGSLS
ERIMQKYGDTIEGMVASAIEYIAVCEKLNYRDVVFSHKSSNPKIMVTAYRQLAKDLDARG
WLYPLHLGVTEAGMGVDGIIKSAVGIGTLLAEGLGDTIRCSLTGCPTTEIPVCDSLLRHT
SIMMAGREEHBETLGETUNDAMTPARTET WEDER OVERLECHTE TPVDBELLCH
ANDERSTERVANTTPREAVVVICTURPARTET WEDER OVERLECHTE TETTPEELLCH
VHQAPPVHFHASDPFIHTSRDFFEKQGHQKPTKLVFSRDFDNKEEAAISIATEFGALLL
DGLGEAVVLDLENLPLQDVLKIAFGTLQNAGVRLVKTEYISCPMGRTLFDLEEVTTRIR
KRTOHLPGLKIAIMGGIVNGPGEMADADFGFVGSKTGMIDLYVKHTCVKAHIPMEDAEEE LIRLLQEHGVWKDPEETKLTV

CPn_0374 420209 420961
CT056 hypothetical protein
vDSMTLSFHTHPLNYWTFEFFDGLPIRHGVFSKQKDAEGTVFAAKNPEIASALQSPKYCD
LHQRHGTSVRCVTPTSFTYQFADGLCTOSPLLSLHIRHSDCQAAIFYDREHHAIANVHSG
WRGLLGNIYAVTVGTMKKLFHTKPQDLFVAIGPSIGPDYAIYPDYATLFPRSFLPFMNPK NHFDLRAIARKQLTNLGISKDRIFISDLCTYTEHDAFFSSRYLAHHPDPNLTGQHSKNRN NVTAVLLLPRD

CPn_0375 421112 421615 No robust homolog present in Genebank/EMBL as of 11/7/98 RLSMKLGASTNHKVHEPVKPKKAQLAEIEANKTQATEGTLRSKSLALQIARAVLYILFAA LMLAAGIFYVTLALGFPLIQAYSIAGIITLVGLAIGLVLLILSLLPKEDEEADALSRNA LLPLTIIVIEQQPITPKPEIPYSYLTKLALLTSLFLTLRRSSSQRKTH

CPn_0376 421680 422294
No robust homolog present in Genebank/EMBL as of 11/7/98
FKVVTAKAPNLTEIRDHGARVPSLFLLSPETSHWKGDKEVSAPLKQLQDLLGEEQWEAMK
TKMNSRKKAGQWALFNSPTPGVSSTLVLAWTPWGYYDKDVQDILERKDPWSSSLSEKDSK EFLKNLFVDLLENGFTSVHIHAEEAFTPLDHTGKPHFKRDNVYLPGKLLGALNEAAVQAN VSADTQFTLFLTQDECNPFHDKKRG

CPn_0377 423441 422347

CPH_037/
sucB-Dihydrolipoamide Succinyltransferase
IMFTEVRIPNIAESISEVTVASLLVTEGALIQENQGLLEIESDKVNQLIYAPVSGRIFWE
VSEBDVVEVGGVVGKLEPAGEGEELGOSQSKETIEAEIICFPQSGVRQSPERKTFIPLR
DQFDOGSGOLSAGDRGFTERMTSTRKTISRFILISAHLESAMLTTFNEVYMTELFHLIRKE
KQEEFLSRYGVKLGFMSFFVKAVLEALKAYPRVNAYIDGEEIVYRHYYDISIAVGIDRGL
VYEYIRDCDKLSNGEIEQKLADLALRAREGLLAIAELEGGGFTITNGGVYGSLLSTPIIN
PPOYGILGMHKIEKRPVVLDNEIVIADMMYVALSYDHRLIDGKEAVGFLVKVKEGLENPA
SELDL

CEA_0378 426195 423445
sucA-Oxoglutarate Dehydrogenase
IVF1EFNYFMDSEFVGQVYSSDMDWIESMYQRFMNHETLDPSWKYFFEGYQLGQAASPSE
ASTKISGNETIAMLQEQKSQFLCTIYFYGYLQSQISTLAPFTDSRFIQEKIAKIDLDEQ
VEŠÄGLLPKAQVSVRELIEALKKCYCGSLTLEFLTCTPELQEFVMNLMEKRQVERFAEQL
LÆSYKDLCKATFFEEFLQIKFTGQKRFSLEGGETLVFMLEHLVHYGSALGISNYVLGMAH
RGRENVLTNVLGKPYRYVFMEFEDDPAARGLESVGDVKYHKGYVLKSHQKDRETTFYMLP RGREWLITWILGKPYRYYFMEFEDDPAARGLESVGUVKYHKGYVLKSHOKDRETTFVMLP
NASHLESVDPIVEGVVAALQHQGHAGKEQSSLAILVHGDAAFSQGVVYETLQLSRVPGY
STECTLHIVVNNYIGFTAVPRESRSTPYCTDIAKMLGIPVFRVNSEDVVACIEAIEVALQ
VREAFSCDVIIDLCCYRKYGHNESDDPSVTAPLLYDQIKKRKSIRELFRQYLLBQQFADI
SECTLASIEKEIQESLNREFQVLKGTDPEPFPKKECHHCDRLNNGELIHHDCOVSLDRET
LEHMSSRLCGFPDNFHHPPKIKTLLEKRMMABEGGVCYDWAMAEELAFASLLIEGVNLRL
SQQDSIRGTFSQRHLWNSDTVTGDTYSPLYHLSAEQGSVEMYNSPLSEYAILGFEYGYAQ
QAERTLVLWEAQFGDFANGAQIIFDQYISSGIQKWDLHSDLVLLPHGYEQGOPEHSSRR
LERYLQLAANNMFQVVLPSTPVQYFRILREHAKRDLSLPLVIFTFKLLLRYPQCVSIEE
FTEEGGFRAILEDADPNYDASILVLCSGKIYYDYAEMLPQDRRKDFSCLRIESLYPLALE
DLYSLIDKYSHLKHFVWLQEESKNMGAYDYMFMALQDILPEKLLYIGRPRSSSTASGSAK
LSR@ELVTCMETLFSLR LSROELVTCMETLFSLR

CPH-0379

CPITO379 426268 426765
CT053 hypothetical protein
KNKKMLCTCSRIODGNPWMKSERLKKLESELHDLTQWMQLGLVPKKEISRHQEEIRILEH
KIYEEKERLQLLKENGEIEEYVTPRRSPAKTVYPDGPSMSDIEFVEPTETEIDIDPGETV
ELELTDEGREDGAVEVDYSHEDDEDPFSDRNRWRRGGIIDPDANEW

CPn_0380 426671 427876
hemN-Coproporphyrinogen III Oxidase
KSTIPTKTMKTLSAIAIAGDAVVSLIPPMLMNGKAPLALYIHIPFCTKKCRYCSFYTIPYK
SESVSLYCNAVIQEGERKLAPIQETHFIETVFFGGGTPSLVSPLDLKRILKELAPHAREI
TLEANPENLTVSYLRQLQETPINRISVGVQTFDDSILQLLGRTHSSSAAITALQECQNHG
FSNLSIDLIYGLPTQSLEIFLSDLHQALTLPITHISLYNLTIDPHTSFYKHRKILVPTIA
QEEILAEMSLLAENLLLSQGFQRVELASYAKPDYPAKHNLYWHTDRFFLGLGVSASGYLH
GERSKNYSHISHYLRAVRKNLPTQETSEILPKKERIKEALALRLRLLEGADLAEFPSTLI
SMLTODVKLQNLFSVHGQCLALNRQGRLFHDTIAEEIMGYSF

429836 428037

CPn_0381 429836 428037

CT326 similarity
SLPNKFRALMTAPPESRSSPPTLLEETEPLSPNPIPADIQIPRITISPPSLDVSTVASSA
EDISVFIAGGPRSSSSAVASDVYELVCLCGGDEDPEPPDSEVRTLYVNGSWOTHOEAVQ
ELL/ISEVRGEAVRLLYNDGSGMSFWPISPCRTLPTLDHPLCOALLTWEOFFSAPENQN
REFLVIFYGDASPY 1QOALTOSRHSPRIVVYGISPTVFIQGDFRVHNYRVSGDFFSSLDC
RGTPAENTTILPYSSGLEGVFLPSIRCPSFTWAVRFGEQCLVANRGEDVEDRGGLSQDAE
RSQLPHGERDLAVVIDSTDFSSMSRLVEWLNGCSPSSDMEINPYPORCPDVALSALYAIS
RYGGLAGOBYTLASVHEGLDLDICYSLILMHTTPATKYFFLLFTNYPQGREPRTARIVAQ
ELVEBILVLVFDCGNVLRKLWMPOEILRAIFISACTISGSIVFVECTRWMGRGLRHRVQ
OFVORVIGGGLEVGTVRASYRDRAGFILGFLQTVHGGLYLPVSIMVLNQIAIQVPRILV
PHHTTAVYDLHNKSAEENWSSGDVLAVQQTLNFILCAFVLFVNLWFFKKSVLRHSRRRR

CPn_0382 430752 430036
yalx/ytal/SAM-Dependent Methytransterase
PVTLYLLENTIXTRAVETLESVICEEVIRLDCLIVESDROGRAFLSLWKIPEVHKEPLAI
SKHARLIKAMDPYLEDFIVKHGENWGLISDAVIDCIADPGASLVRRARALGIPVQAFSGP
CSITLALMLSGLPSGSFTFLAYLPGSPKERVKSIKKAATSKEVSTSVCIETS/RNVYTFE
SLLDFLPSYAELCVASDL/GPSELVLTROVQSWRTTEDLGSVKOSITKVPTIFLFHIPN

CPn_0383 431711 430749
CT047 hypothetical protein
VQDTTFUTLPHQKSLTSFDDFSQAYAEKVPAIALIGSALEDDKDALIELLVGESFKELGG
QGLMPATLMSWTETFALFQEHETLGIIHAEKFPLATKEFLSRYARNPQPHLTILIFTTKQ
ECFRELSKALPSALSLSLFGEWPADRQKRIIRLLLQRAERVGISGSGSLAGLFLRALAST
SLPDILSEFDKLLCSVGKKTSLDHSDIKELVVKKEKASLWKFRDSLLKRDPVEGHQQLHF
LLEGGEDPLGIITFLRTQCLYGLRSIEEGSKENKHRMFVLYGKERLHQALNSLFYAETLI KNNVODPIVAVETLVIRMVNL

1 11 (1 nctB-Histone-like Protein 3

NGCB-HISCONE-TIKE PIGLEIN J VITCLIRGIKMIGAQKKQSGKKTASRAVRKPAKKVAAKRTVKKATVRKTAVKKPAVRKTA AKKTVAKKTTAKRTVAKTVAKKPAVKKVAAKRVVKKTVAKKTTAKRAVRKTVAKKPVARK TTVAKGSPKKAAACALACHKNHKHTSSCKRVCSSTATRKHGSKSRVRTAHGWRHQLIKMM SR

CPn_0385 434042 432522
pepA-Leucyl Aminopeptidase A
FLVIKGEFVVLFHAQASGRRRVKADAIVLPFWHFKDAKNAASFEAEFEPSYLPALENFQG
KTGEIELLYSSPKAKEKRIVLLGLGKNEELTSDVVFQTYATLTRVLRKAKCSTVNIILPT
ISELRLSAEEFLVGLSSGILSLAYDYPRYNRVDRNLETPLSKVTVIGIVPKWADAIFRKE
AAIFEGVYLTRDLVNRNADEITPKKLAEVALNLGKEFPSIDTKVLGKDAIAKEKMGLLLA VSKGSCVDPHFIVVRYQGRPKSKDHTVLIGKGVTFDSGGLDLKPGKSMLTMKEDMAGGAT VLGILSALAVLELPINVTGIIPATENAIDGASYKMGDVYVGMSGLSVEICSTDAEGRLIL ADAITYALKYCKPTRIIDFATLTGAMVVSLGEEVAGFFSNNDVLAEDLLEASAETSEPLW RLPLVKKYDKTLHSDIADMKNLGSNRAGAITAALFLQRFLEESSVAWAHLDIAGTAYHEK EEDRYPKYASGFGVRSILYYLENSLSK

CPn_0386 434543 434046 ssb-ss DNA Binding Protein KSKGYLMMFGHFAGYLGADPEERNTSKGKRVITLRLGVXTRVGMKDETVWCKCNIWHNRY DKMLPYLKKGSGVIVAGDISVESYMSKDGSPQSSLVISVDSLKFSPFGRNEGSRSPSLED NHOOVGYESVSVGFEGEALDAEAIKDKDMYAGYGOEOOYVCEDVPF

435229 434699

CT043 hypothetical protein
NNNNLLQGDSLMSRQNAEERLKNPAKELKLPDVAFDQNNTCILFVDGEFSLHLTYEEHSD
RLYVYAPLLDGLPDNTQRKLALYEKLLBGSMLGGQMAGGGVGVATKEQLILMHCVLDMKY
AETNLLKAFAQLFIETVVKWRTVCADICAGREFSVDTMFQMPQGGGMQPPPTGIRA

CPn_0388 435323 437320
glgX-Glycogen Hydrolase (debranching)
STMEKVSSYPSVPLPIGASKISPNRYRFALYASQATEVILALTDENSEVIEVPLYPDTHR
TGAIWHISIEGISOSSYAFRVHOPKKHGMOVSFKEYLADPYAKNIHSFQSFGSRKKQGD
YAFCYLKEEPFPWDGDQPLHLPKEEMIIYEMHVRSFTQSSSSRVHAPGTFLGIIEKIDHL
HKLGINAVELLPIFEFDETAHPFRNSKFPYLCNYWGYAPLNFFSPCRRYAYASDPCAPSR
EFKTLVKTLHQEGIEVILDVVFNHTGLQGTTCSLPWIDTPSYYILDQGHFTMYSGGGNT
LNTNRAPTTOWILDILRYWYEEMHVDGFRFDLASVFSRGPSGSPLQFAPVLEAISFDPLL
SCTUILA EBWINACOLVWINVEEMI SDBWSEBWINDDENWASH WATGHT LOTTAGED ISOS ASTKIIAEFWDAGGLYQVGYFPTLSPRWSEWNGPYRDNVKAFLNGDONLIGTFASRISGS QDIYPHGSPTNSINYVSCHDGFTLCDTVTYNHKHNEANGEDNRDGTDANYSYNFGTEGKT EDPGILEVRERQLRNFFLTLMVSQGIPMIQSGDEYAHTAEGNNNRWALDSNANYFLWDQL TAKPTLMHFLCDLIAFRKKYKTLFNRGFLSNKEISWVDAMGNPMTWRPGNFLAFKIKSPK AHVYVAFHVGAQDQLATLPKASSNFLPYQIVAESQQGFVPQNVATPTVSLQPHTTLIAIS

CPn_0389 438254 437319
CT041 hypothetical protein
TVFNFKRFYQKDSQRQNSNTTCLRFFKKTCKELIEFRRRTVKLLKNVLLGLFFSMSISGF
SEVKVSDTFVKQDTVVEPKIRVLLSNESTTALIEAKGPYRIYGDNVLLDTAIQGQRCVWH
ALYEGIRWGEFYFGLOCLKIEPVDDTASLFFNSIQYQGSLYVHRKDNHCIMVSNEVTIED
YLKSVLSIKYLEELDKEALSACIILERTALYEKLLARNPQNFWHVKAEEEGYAGFGVTKQ
FYGVEEATDWTARLVVDSPQGLIIDAQGLLQSNVDRLAIEGFNARQILEKFYKDVDFVVI

CPn_0390 439171 438134
ruvB-Holliday Junction Helicase
RKSDREGSYMTHQVAVLHQDKKFDVSLRPKGLEEFYGQHHLKERLDLFLCAALQRGEVPG
HCLFFGPPGLGKTSLAHIVAYTVGKGLVLASGPQLIKPSDLLGLLTSLQEGDVFFIDEIH
RMGKVAEEYLYSAMEDFKVDITIDSGPGARSVRVDLAPFTLVGATTRSGMLSEPLRARFA
FSARLSYYSDQDLKEILVRSSHLLGIEADSSALLEIAKRSRGTPRLANHLLRWVRDPAQI
REGNCINGDVAEKALAMLLIDDWGLNEIDIKLLTTIIDYYQGGPVGIKTLSVAVGEDIKT
LEDVYEPFLILKGFIKKTPRGRMVTQLAYDHLKRHAKNLLSLGEGQ

CPn_0391 439701 439510
No robust homolog present in Genebank/EMBL as of 11/7/98
KDDLYKQEKPIPKATILSRNLEVMLDNPKGKRQTLFLGRTSGRSALYSYSRRILVLLNAF

CPn_0392 439814 440383 dcd-dCTP Deaminase

MSIKEDKWIREMALNADMIHPFVNGQVNVNEETGEKLISYGLSSYGYDLRLSREFKVFTN VYNSVVDPKCFTEDIFISITDDVCIVPPNSFALARSVEYFRIPRNVLTMCIGKSTYARCG IIVNVTPFEPEWEGHVTIEISNTTPLPAKIYANEGIAQVLFFESSTTCEVSYADRKGKYQ KQQGITVPCV

CPn_0393 440229 440723
CT038 hypothetical protein
KFLTLRHCQRKFTLMKGLFRSYSLSLVPPARFLMQTEKESIKSNKASPYLVSKVSVRKKN
WGFRLLEEVMIKSMWVIFSILICOFFV/DRAIOELRTEELRLQSKVSSLCQDILSAQEKQR
QLQLHLQHWQDSAAIEAALIORLGLIPYGYKKLCVSPKQQSENKD

CPD_0194 440727 441958
ElyC=CBS Domain protein(HemolySin Homolog)
KETMIPTMIMPETICPTICOGFIOLOGIALPOLIPTMICHYKEKSKKOORVATLLEHPH
HELITLIPCDIGENIAJOMERA LEFGEAASWMETWILPLAITHINGETEPKAVALPFNTO
EASSWAPELIAWTKIEKPELHWGIVGIHYVWWAILERGO DEI IQU-QELKEVLQSKEORGV
VNQEESPLLYGYLSESOCSVKERMGPPGDILFYDIOTPHENLYLLESKOHGSRVPICNDN
LQNLUGICTARSILLBIKKPLGGSDDLIFLEKPPFYMMETTISAKMALCQMAAEDETISMII
DEYGGIEGOFTOEDEFETWGE LYDGHAKKILFTESADVILAGSTLELREFSEIFDINL
PPNNMIATICKWGIEQIGTIPTTOMKLSWMESEPDULDAAENRIRKVYIRKLYD

CPn_0395 441955 443175
CT257 hypothetical protein
GNCMTNSALFWIGOVIICIVLQGFYSMMEMACVSFNRVRLQYYLTKDHKKARYINFLIRR
PYRLFGTVMLGVNIALQVGSESSRNCYRALGITPDYAPFTQIFIVVIFAELLPLTISRKI
PEKLALMGAPILYYSHYIFYPLIQLIGSLTEGL/YLLNIRKEKLNSTLSRDEFQKALETH
HEEQDFNTIATNIFSLSATCADQVCQPLEQVTMLPSSANVKDFCRTIKNTDINFIPVYHK ARKNVIGIAHPROFVNKALDEPLINNLHSPWFITAKSKLIRILKEFRONRSSVAVVINAS GEPIGILSLNAIFKILFNTTNIAHLKPKTISVIERTFPGNSRIKDLQKELDIQFPQYPVE TLAQUADEBLERITATVITSVITINLLE SATTATLYGIKTVGIKNIL

CPn_0396 444J59 443241
yhfo-nifs-related protein
ysmiyldnnamippergllefloktfliegtyanpssvholgkksrolvleashwmokvl
srogrvlytsgateslalaiaslpkdshvitsgsehpaileplkhsslsvsylnpeegrc
vltieoieravipktsaiildavnsetgakadiaaiahfaoerolofivdaranvokeri
vlpsgvtmaafsghkfhalsgigallvspgvklhpolwogogogograftenlwgiasll
yifkyldlhoerisoeilithrngfekaikaripdvhihcadoprannvsaiafpplegev
Loialdiegvacgygsacssgatapfkslvsmgvdeeltlatlrfsfshllloedverav
giiekvverlkns

444381 CPn .0397 445124

CPn_0397 445124 444381
PP2C phosphatase famlly
EHFVDFDYFGLSDIGRVRARNEDFWQVNLMSQVVAIADGVGGRLGGDIASQEAVTSLMEL
IDEQQSKLMGYGDDQYKETLKKILLEVKGVVYEHGMEEHLQGMGTTLSFIQFRKDRAWL
FHVGDSRIYRIREGELRRIFEDHSLENQLKNRYGLPKQSDKVYSVRHILTINVLGSRPVM
PDIRNLPCEKEDLYCLCSDGLTNMVPDIDIRDILNQPATLEERGNALISLANTRGGDDNA

CPn_0398 445518 445700
No robust homolog present in Genebank/EMBL as of 11/7/98
IEEELPMQIENSSILFAEVVMKWFIFSVISAPVVFLPGCTLIPKEKVTKVPSQLWSESLSQ

CPn_0399 445759 446523
CT253 hypothetical protein
YKLMRVLMSKSLNCESIDLKSKNFFRARIFCKISNLRTVTMRKMLVLLASLGLLSPTLSS
CTHLGSSGSYHPKLYTSGSKTKGVIAMLPVFHRPGKSLEPLPWNLQGEFTEEISKRFYAS
EKVPLIKHNASPOTVSQFYAPIANRLPETIIEOFLPAEFIVATELLEQKTGKEAGVDSVT
ASVBVRVPDIRHHKIALIYQEIIECSQPLTTLVNDYHRYGMNSKHFDSTPMGLMHSRLFR EVVARVEGYVCANYS

CENTIO400 446527 447306
CT254 hypothetical protein
SKEMSKFILLLSLGVAALASKNFFIMPAPSGKTPLKLRQVLFGGALLVFSSLVALSVSSQ
TAELLSTMTGISLAFAPLFYLLFLPKDITRAILFSGERPVKTSWRALGSAIRMWIIIIPV
TGEFGIMMSKFLTLVLPTQEHTQEVTQEVQNSLPITGHYISMILNIGVLTPFGEEVFFR
GILGTFLKNKMTRIAAVLCSSIIFSFIHIEHSLGSWVFVPVLFVFSLSAGFLYEKDRHIL
SPIFALHGLFNLTSLLFLGIK

447884 447495

CF255 hypothetical protein
MRDHAFSKLIGTVRAMVVEGRCPWSLQQSLVSMVEHILGECQEFHEAVLQGKTVQEVGSE
AGDUTLVLILCFLLEREGVLASEDVANEAMEKLRRAPYIFAEDYKPVSIEEADRLWEL

CPN_0402 449012 447888
mitY-Adenine Glycosylase
mitY-Adenine Glycosylase
mekRFCMTKIAFSEKAKNFPVEALKKWFEKNKRSLPWRDNPTPYSVWVSEVMLQQTRAEV
VEDYFNQMMERFPTIESLAAAKEEDVIKLWEGLGYYSRARHLLEGARMYMEFHGKIPDD
AESLAQIRGVGPYTVHAILAFAFKRRAAAVEDOWLRVLSTIFLIEFSIDLESTRTWYSRI
AGALLPHKSPEVIAEALIELGACICKKVPQCHRCPVRQACGAWRENKQFVLPVRHARKKV
IFILHRLVAIVLYGGSLVVEKRPKEMMAGLYEFPYIEVEPEEGLQDIEGFTKKMELSLES
DIEGETGIN KOOBUB GENTWEINLED. PLEFLGNLKEQRHAFTNHKVHLCPIIFKATSLPQFGELHLLSDIDHLAFSSGHKKIKDAL LIXEGDVRSRESIGV

CPE_0403 449009 449710
ycc_-predicted pseudouridine synthetase family
nFMGLSNDKRAALQYFMENFSWLATQVSRLSSFLRSQLPNHSKQEILASIRQHRCRVNGF
LERFESYKVQPGDRVSLSLIPSTKQQPSILWEDDYSIIYEKPPHLTTEQMAHMTRFFTVH
RLDKGTSGCLLNGKSKQAATELMKLFKQRKIHKQYIAFVFGHPKKKFGTVKSYTAPVYRR
CGAVIFGAAGPSQGEPIKSAYKWDCWVILLSEMSTTDLKNSLPRSSALSSMLTP

CPn_0404 450962 449871
No robust homolog present in Genebank/EMBL as of 11/7/98
ELEALEQKYGKAVLLIALSELGIDTMSLLSGHRLEGFPPIAEVMAACDRCSMDFCEILKS
QSMDLWADAASCVDGLLQDPFWSTAIASGIAKSSLOETEFECESKVMVLSSWGEQGAQVC
SPFNLERICMSFPSLKVYSLKKNGCENMGIQLSASCMNLLMSIFFVATMGGSTPINITKE
NLMALVALVLSHYOCYFVPATGDPORGNILGNPEVNAILARGMGMRVDLERKRGGESSSS
RYLELAARCFENSLTKTSLLSDANNVQERDKCLLQMSTSLMHTAGLNLQRPPVPTPSGVT
AHPOPQPDPVVTSQPSLLGARERSPVSSRGRFPVVLPLSVISPRSHPGRVERRDLEDEEE
EVMF

CPn_0405 451814 450966
CT105 hypothetical protein
NIQTSHSRVLLKKFSKEFTIRTYRSLGFTDYLGGCLTNPLGKFPSPQNPQVVTIAPSSTT
PQAVSSAVOGFLQTGGAASSTATTTTASGASALGLSPDQVQALLINLLINVGQPSVGQPST
SAGTSGASSSSASMQQQLLQLILDRTTGSGGSSVSSEQLQQLLSLUSQNTTSQGGSGGTQ
AGQAASVLLNLLSATGSAAANPLGTAASLAQIIYAAVTSPGAKKTSEFCYNYCGETCQGN
UGCPTCGCPDGQCGCGGFGRFFCGVWKNCCGIGEGSQEPAIPL

452855 451960

CETT_UTUS 4 13.1304 43.2305 tubf-Enoyl-Acyl-Carrier Profein Reductase CGFMLKIDLTUKVAFVAGIGDDQQYGWGIAKLLNEAGATIIVGTWVFIYKIFSQSWELGK FNEJRKIZNOTILLETAK LYPMDAJPDJEPDVPEDIAENKRYKO LYGIPT JEVAROVIKOP GHI DI LIVIJILANGPE DIKIJLETARKOYLAALISAJI ZEPVSLLJIHOS IMBIRKIST I DLIV YLAIMRAND POTYKRIMISTAKAALESDPKYLJANEAJRENG I RVATI LISADPLASRAKKA IGE I ERMVDYYQEWAP LIEAMNA BOVGAVAAFLAGPLAJA [TGETLYVDHGANVMG I GPEMPPK

CPn_0407 453757 452858 HAD supertamily hydrolase/phosphatise

NYGDAMEKLLVTDID/TTTTP/DSHHLDKKV/ERLYALHQAGWKLFFLTGR/YKYAAPLFSD FDAPYLLCCQNGASVWGGTUSNLLYSKSLP/SDLLCILQDCMEGATALFSVESGAPYGDHY YRFSPTPIAQDLHEYVDRR/FPNAKEREILFETRSLKDDYAFPSFAAAKVFGLRDEVIRI QKELERQEALTSVATMTLMRWPFDFRYAILFLTDKSVSKCKALDRVVNILYDCKKPFVMA SGDDANDLDLIERGDFKIVMSSAPEEMHVHADFLAPPADRNGILGAWEAGVRYYDDLMSL

454090

CPn_0408 454090 454581
CT102 hypothetical procein
spvLimings: The transfer transfer the transfer of the transfer transfer the transfer of the EFPPDTDINHLLQENIIKQSSSSATSDFLSYAICKIQKKQVTNQL

CPn_0409 454645 455127
CT260 hypothetical protein
MTTWTLNQNNLTKFLKSSDEEPFLERESGLTYINIQANGNELPLFFVIRSEGEILQLICY
LPYQLHESHKASTAELHLLNRDIDIRGFGMDEEQGLIFYRLVLPCLNGEIHDTLLRIYI
DTIKLVCDSFSHAIGLISSGNMNLDELRRQALQEQQEKRNE

455087 455833

CPn_0410 455087 455833
dnaQ-DNA Pol III Epsilon Chain
DVRLFKSNKKNVMSSQTMDVLIFYDTETTGTQIERDRIIEIAAYNSVTDESFLTYVNPEI
PIPDEASKIHGITTDAVLSAPKFPEAVEGFRKFCGEDSILVAHNNDGFDFPLLGKECRH
SLEPLTNRTIDSLKWAQKYRPDLPKHNLQYLSKQVYGFAENQAHRALDDVVILHKVTSLI
GDLPPQQVLDLLQQSYHPKVFKMPFGKYKGQPLVDIPKSYFEWLENQGALDKPENKDIKA

CPn_0411 455794 456609
CT262 hypothetical protein
RHOSRYSSITSTDMILTAAFSPCPNDIFLERSFLKDPQFRPLLNQVTIADIETLNTLALQ
RRLSLMKMSAALFPLVSDYYNLMDVGNTLGYNSGPIVLSLDPECSLDTLATPGEMTTAHA
LCKLYYPKAKLIPMFYDKILSAILQGKVDGGALIHEERFSYDLQLTLRADFGELWRRKTI
FPLBLGCLAIARYVPMATVDALTAALRKSLICSLKDPITAGAKAVEYSKNKNVTVIHRFI

GTYINKETFQLSKTGKKALHMLWKANECCQYT 456515

457246 CT263 hypothetical protein
EPISTKKPFMYLKLGKKLYICSGRPMNAVNTPKKILCIVADYREISPLIEQLDFTQINEH LFISIKAPFNYLKLGKKLYICSGRPMNAVNTPKKILCIVADYREISPLIEQLDFTQINEH LYSYRCTDYHLDLYIVHVWGSTAVLNALQSYCQAYTDYDLWINPGFVGACSPEIPLGQCY TIEKIANLTTDTPPVLSEDPPYIFDALPDSLFKSSLVTSPVLYHYGFHKTFKLLDMBGYA IASQAAEHHIPCSFLKITSDYTVFGDCPFSRLEEVSQKLTQTLVELLPELMERAIPPKLL LPCP

CPn_0413 459209 457227
msbA-Transport ATP Binding Protein
VFMKLLKAVLRHKNHLVILGCSLLAILGLTFSSQMEIFSLGMIAKTGPDAFLLFGRKES
GKLVKVSELSQKDILENWQAISKOSETLTVSDATTVIAEHGKSTASLTSKLSKFVRNYID
VSRFRGLAIFLICVAIFKAVTLFFQRFLGQVVAIRVSRDLRQDVFKALQQLPMTFFHDHD
IGNLSNRVMTDSASIALAVNSLMINYIQAPITFILTLGVCLSISWKFSILICVAFFFIRI
IJVVIARRIKNLAKRIQKSQDSFSSVLYDFLAGVMTVKVPTEKFAFFKYCEHNRIFISL
EEKSAAYGLLPRPLLHTIASLFFAFVVVIGIYKFATPPEELIVVCGLLYLIVDPIKKFGD EERSAAYGLERPLEHTIASLFFAFVVIGHTFA AFPEELTYGLEITHITPIAATGE
ENTSIMGCAAAERFYEVLINHPDLHSQKERETEFLGLSNTITFENVSFGYGENHILMNL
SFTLHKGEALGIVGPTGSGKTTLVKLLPRLYEVSQCKILIDSLPITEYNKGSLRNHIACV
LQNPFLFYDTVMNNLTCGKDMEEEAVLEALKRAYADEFILKLPKGVHSVLEESGKNLSGG
QQRLAIARALLKNASILILDEATSALDAISENYIKNIIGELKGCCTQIIJAHKITTLEH
VDRVLYIENGQKIAEGTKEELLQTCPEFLKMWELSGTKEYNRVFVPDHKLVANPTDMAIT

CPn_0414 460203 459172
accA-AcCoA Carboxylase/Transferase Alpha
LCLRIVCIKMILFIRGHILMELLPHERQVVEYEKAIAEFKEKNKKNSLLSSSEIQKLEK
RLDKLKEKIYSDLTPWERVQICRHPSRPRTVNYIEGMCEEFVELCGDRTFRDDPAVVGGF
VKIQOQRFVLIGQEKGCDTASRLHRNFGMLCPEGFKKALRLGKLAEKFGLPVVFLVDTPG
AYPGLTAEERGQGWAIAKNLFELSRLATPVIIVVIGEGCSGGALGMAVGDSVAMLEHSYY
SVISPEGCASILWKDPKKNSEAASMLKWHGENLKOFGIIDTVIKEPIGGAHHDPALVYSN VREFIIQEWLRLKDLAIEELLEKRYEKFRSIGLYETTSESGPEA

CPn_0415 461522 460221
CT266 hypothetical protein
SOTsflecLtItFVIIIWMCNAFLIKLCVIMGLQSRLQHCIEVSQNSNFDSQVKQFIYAC
QDKTLRQSVLKIFRYHPLLKIHDIARAVYLLMALEEGEDLGLSFLNVQQYPSGAVELFSC
GGFPMKGLPYPAEHAEFGLLLQIAEFYEESQAYVSKMSHFQQALFDHQGSVFPSLWSQE
NSRLLKEKTTLSQSFLFQLGMQIHPEYSLEDPALGFWMQPTRSSSAFVAASGCQSSLGAY
SGDVGVIAYGPCSGDISDCYYFGCCGIAKEFVCQKSHQTTEISFLTSTGKPHPRNTGFS
YLRDSYVHLPIRCKITISDKQYRVHAALAEATSAMTFSIFCKGRNCQVVDGPRLRSGSLD sykgpgndimilgendainivsaspymeifalogkekfwnadflinipykeegvmlifek

KVTSEKGRFFTKMN

CPn_0416 461871 461557 himD/ihfa-Integration Host Factor Alpha EALSNMATMTKKKLISTISQDHKIHPHHVRTVIONFLDKMTDALVKGDRLEFRDFGVLQV VERKPKVGRNPKNAAVPIHIPARRAVKFTPGKRMKRLIETPNKHS

463047 462244

CPT_U41/ 45344/ 45344/ 45344 and ase amia-N-Acetylmuramoyl Alanine Amidase REKOMKLTKYLNTKQLRSMISRLFVRYSLPMSKQLSFFALCVLGSHPIFAQTPNPPQRVR RSEVIFIDPGHGKDQGTASKELHYEEKSLTLSLALTVQSYLKRMGYKPQLTRSSDVYVD LGKRVALGNRGQXDVFISTHCNHSSNAAAFGTEVYFYNGKYGSPTRNMSSVLGKNILAA MEKKGILKSRGKKTANFVVIRDTSMPAVLVETGFLSNSFERAALQDARYRMHVAKGIAEG VHNFLCGPSFQKPKQNIAKIRKPQIQAN

CPn_0418 464401 462953
murE N Acery Imuramoy Ialany Iglutamy LDAP Elgage
MDLKELLIKOVOAK LYCKVROLEVRNITTROSKOVSVODIF LAHKOOP (IXMDFAVDALANG
ATALAJOLYNPELSVOJ CTPNLEELEAELSAKYYEYPOSKLHTT IGJYTSTNYFTTVTOLI
KALLDJYOKINS JAUT EHHILGEGVI POSTTTPTPALLOF/LATMYPONROAJVMEVGSI
GLAKSRVAYTNI UDAVLEN LELGEGY PHOSTEDTYVAAKAPLESUVERSMVY INTOSPYA
DR TEJAKAPV TYGG ENAADYPATD IGLASGOTKYTLJYGDQK LATDSE IGKYNVYNL
LAAT TJYHA HECHILGILE KIGLYG PHOSTLDPVLMGPCEVY LDYAHPPDALLANVINGL
HELLE FYSRLIVVEG A APROPORPRILAQVVERYGFAJJTSDNERGEPED LINEIGDG
FY SRYFFIG IDRKQALTYALSHASDPDIVLIAGKGHEAYGIFKHOTJAEDDPJTVCEVLA
JV

CPn_0419 466897 464876

pbp3- transqlycolase/transpeptidase
OLFPNTNIWNIPOKKVSVFYPMSYRKRSTLIVLGVFALYALLVLRYYKIOICEGDHWAAE QLFPNTNIWNIEQKKYSYYPMSYKKASTLIVUZSYALYALLVLRYYKIQICEGDHWAAE
ALGOHEFCVRDPFREGTFFANTTVRKGDKDLQOPFAVDITKFHLCADPLAIPECHROEII
OGILGFIEGGTYDDLSLKLDKKSRYCKLYPLLDVSVHDRLSLWMKGYATKHRLPTNALFF
ITDVQRSYPFGKLLGQVLHTLREIKDEKTGKAFPTGGMEAYFNHILEGDVGERKLLRSPL
INRLDTNRVIKLPKDGSDIYLTINPVIOTIAEEELERGVLEAKAQOGRLILMNSQTGEILA
LAQVEPEDHERDIGFNUNGFFIEGERGESTEN SYNTERGIMENTETTAT LEGANEEASLKSC
KKIFDPERDIFTUNGFFIEGERFFERFESTEN SYNTERGIMENTETTAT LEGANEEASLKSC VAWYQQKLLALGFGRKTGIELPSEASGLVPSPHRFHINGSLEWSLSTPYSLAMGYNILAT GIOMVOAYAILANGGYAVRPTLVKKIVSASGEEYHLPTKEKTRLFSEEITREVVRAMRFT TLPGGSGFRASPKHHSSAGKTGTTEKMIHGKYDKRRHIASFIGFTPVESSEGNFPPLVML VSIDDPEYGLRADGTKNYMGGRCAAPIFSRVADRTLLYLGILPDKKLRNCDEEAAALKRL YEEWNRSPKQGGTR

CPn_0420 467120 466824 CT271 hypothetical protein KSFFMNKSRFLRLCCCLCFCGSLFYFYINKQNSLTKLRLEIPCLSVRLRQLEQQNISLRF LIDKIERPDHLMEIAALPEYQYLEYPSEESISLLSYELP

468007

CPH_0421
yabC-PBP2B Family methyltransferase
EILMSERAHIPVLVEECLALFAQRPPQTFRDVTLGAGGHAYAFLEAYPSLTCYDGSDRDL
QALAIAEKRLEFFQDRVSFSHASFEDLANQPTPRLYDGVLADLGVSSMQLDTLSRGFSFQ
GEKEELDMRMDQTQELSASDVLNSLKEEELGRIFREYGEEPQWKSAAKAVVHFRKHKKIL
SIQDVKEALLGVFPHYRFHRKIHPLTLIFQALRVYVMGEDRQLKSLLTSAISWLAPQGRL
VIISFCSSEDRPVKWFFKEAEASGLGKVITKKVIQPTYQEVRRNPRSRSAKLRCFEKASQ

CPn_0422 468233 468784

CPT_U422 400233 400764
CT273 hypothetical protein
GLAWYEIFNYSTSIYECHASNNRIVSDFRKEICMEGISIRDVAKHAQILDMNPKPSALTS
LLQTNQKSHWACFSPPNNFYKQRFSTPYLAPSLGSPDQQDEDIEKISSFLKVLTRGKFSY
RSQITPFLSYKDKEEEEDEDPEEDDDDPRVQQKVLLKALDLGVKSTNVMIDYVISRIFQ

CPn_0423 468788 469216
CTZ74 hypothetical protein
CMLDNEWKAILGWGDDELEELRISGYSFLRQGHYSKAILFFEALVILDPLSIYDHQTLGG
LYLQIGENSQALAVLDQALRNQGDHLPTLLNKTKALFCLGRIEEATAIATYLSSCPIPAI ANDAEALLMSYSKATKKNAALVR

CPM-0424 469528 470961
dnam-replication Initiation Factor
SRCMEIFSPSIMGWODGIWESFINKESGMLTCNECTTWEQFLMYVKTRCSKTAFENWISP
IOVLETQEKIRLEVPNIFVONYLLDNYKRDLCSFVPLDVHGEPALEFVVAEHKKPSAPV
ASDKESNEGISEVFEETKDFELKLNLSYRFDHS IEGPSNOFVKSAAVGIAGKFGRSYNPL
FIEGSVGLGKTHLLHAVGHYVREHHKNLRIHCITTEAFINDLVYHLKSKSVDKMKNFYRS
LDLLLVDDIQFLONRONFEEEFCNTFETLINLSKQIVITTSDKPSQLKLSERIIARMEWG
LVÄHVGIPDLETRVAILOHKAEQKGLLIPNEMAFYIADHTYGNVRQLEGAINKLHAYCRL
FGKSLTETTVRETLKELFRSPTKQKISVETILKSVATVFQVKLNDLKGNSRSKDLVLARQ
IAMKLAKTLITDSLVAIGAAFGKTHSTVLYACKTIEHKLONDETLKRQVNLCKNHIVG

CPB=0425 470965 471564
CT276 hypothetical proteins
FRGCPMFRRTGKGPFEDVQTLYEEETSSPSSYSPYSRSERPETPPSLFDNPKASEARPLN HNTTEESSLPOWSSTPRTESLLPLEEPETTLGEGVTFKGELAFERLLRIDOTFEGILVSK GKIEGPKGVVKADIQLQEAIIEGVVEGNITVSGKVELRGGAIIKGDIQANTLCVDEGVR ILGYLAIAGITDHSERERDL

CP1_0426 47211 471536
CT279 similarity
MVLF3LLFPKLCYGCQAPGAYFCSNCLEKLLVEDREGRCLHCFRYLGSSETRLCSQCSPS
SQEQAFSLYLPSQTALSVYARACEGKRPALQFFSKSIAFELASLDETPSCIAYITSTISR
KIYVEVAKLEKLLRIPLWPWLPKKRQIEKLPKGEGICFLSAYPLSQKWMQTIVGGSASPL

472153 473715

CPRÉGA27

A72153

473715

nqr2-NADH (Ubiquinone) Dehydrogenase
AVCYYFERVEASTFLSTTMLKKFINSLWKLCQQDKYQRFTPIVDAIDTFCYEPIETPSKP
PFIRDSVDVKRWMLVVIALFPATTVAIWNSGLQSIVYSSGNPVLMEQFLHISGFGSYLS
FVYKEIHIVPILWEGLKIFIPLLTISYVVGGTCEVLFAVVRGHKIAEGLLVTGILYPLTL
PPTIPYWMAALGIAFGIVVSKELFGGTGNNILNPALSGRAFLFFTPFARMSGDWWGSNP
GVIKDSLMKMNSSTGKVLIDGFSQSTCLQTLNSTPPSVKRLHVDAIAAMMLHIPHVPTQD
VHSQFSLWTETHPGWVLDNLTLTQLQTFVTAPVAEGLGLLLPTQFDSAYAITDVTYGIG
KFSAGNLFWGNIIGSLGETSTFACLIGAIFLIVTGIASWRTMAAFGIGAFLTGWLFFIS
VLIVQQNGAWAPARFFIPAYRQLFLGGLAFGLVFMATDPVSSPTMKLGKWIYGFFIGFMT
IVIRLINPAYPEGVMLAILLGNVFAPLIDYFAVRKYRKRGV

473719 474681

CPn_0428 473719 474681
"nq3-NADH (Ubiquinone) Oxidoreductase, Gamma"
NMSKGSSKHTVRINGTWYIVSFILGLSLFAGVLLSTIYYVLSPIQEQAATFDRNKQMLLA
AHILDFKGRFQIQEKKEWVPATFDKKTQLLEVATKKVSEVSYPELELYAERFVRPLLITDA
GKVYFSFEEKKLNPIEFFEKYQESPPCQSSPLDFYVILENTSRTENMSCADVAKDLSTVQ
ALIFPISGFGLMGPIHGYLGVKNDGDTVLGTAWYQQGETPGLGANITNPEWQEQFYGKKI
FLØDSSGTTNFATTDLGLEVVKGSVRTTLGDSPKALSAIDGISGATLTCNGVTEAYVOSL
ACYMOLLINESMITHEKKTGE ACYROLL INFSNLTHEKKTGE

CPn_0429 474666 475319

oqr4-NADH (Ubiquinone) Reductase 4
KENPPMTOKKSYKSYFFDPLWSNNQILIAILGICSALAVTTTVOTAITMGIAVSIVTGGS
SFPVSILKRETPDSYRMITQLIIISLEVIVIEQFLKAFFFDISKTLSVEVGLIITNGIVM
GROESLARHVTPIPAFLDGFASGLSYGWVLLVIGVIFELFOFGTLMGFRIIFOFVYASET
HPDSTYONLSHMVLAPSAFFLLSIMIWLVNIRDSKYFYR

CTH_0430 475323 176003

HQL5_MADH (Ubiquinone) Reductable 5

PMMIDAZYMLANYGI LUQAAF [QNI LLANFLOME 37LACCTRYCTANGLOMGVALVLTYT

GOTTMWOTHAF TYCHARALTM GERGLASVNLOFLELT I F I EVVI AAFTQ I LELLLEKVORNLY

LOGGEFLIEL I AVNCA I LOGGEFRSYPF I FMM I FOLDAGGGMIGA EV LAT I KEKLAY

CD I FYBLOMG I GETTYGI, I AMARMOLIG I DI JKE GAK I QRAPLEGEVVENTTNPLKEGS

SKHQPSISKAPTQRRSL

476489 475151

No robust homolog present in Genebank/EMBL as of 11/7/98
KIMTTLPKYVPRSRQNPDTLTFLKRYSSVLLHSENSLSYRIFAKVLAILLTSLAVAFAVT LFSCEGSQLRLCALYIGIALAICVLLTIVVYCIAJKIATACKKPPSISRIEIV

GLFLTGSAPLQLJIWIAAG.ITUSMLVCACWRYKIUNALLKTKVAHES

CPn_0433 477327 476929
gcsH-Glycine Cleavage System H Protein
RTFRILYGTLYRTGSRKVMWYSDYHVWILDVHERVVRLGLTEKMOKNIGAILHVDLPSVG
SLCKEGEVLVILESSKSAIEVLSPVSGEVIDINLDLVDNPQKINEAPEGGGWLAVVRLDQ DWDPSNLSLMDEE

CPn_0434 479471 477276
CT283 hypothetical protein
RPWWRIYQODLFRLCRDFAWFFSLLSFTLRFYCLGRGWTLLSFFYKHQKKFIGIVIAVV
CVSGIGVOWGRFSRKGSAESTSRRTVFTTASGKRYVEKDFMAMKKFFAHEAYPFTGNPRA
WNFINEGLLTDYFLTTRVGEKLFLKVYHPGEKIFSKEKAYQPYRRFDAPFISSEEVWKSS APOLLE ILKVFOOIEMPISKEGFLARAKLFLEERRFPHYVLRQMLEYRQMFALPPDEAL SRGKDLRLFGYGTIODWFGDAYLSAAVELLIRFIDEOKKVLPRPSKQEARDDFYDKAKHA YTKISKMKEFSLGFEEFVNSYFOFLEISESFFNMYRDILLCKRALLLLQGGVSFDFOPL TTFFVQKDSIQVEFFRLPKEYSFKTKQELKAFEVYLKLVSLPKSDSLDVPNEILPIATI THE VOGADS LOVE FIRENESS HIN QUELAFEYS LILLY SEPSIBLID PREILD LATE AKAREPRIVGRAFS IDYKRVALQDLAATVPMVEVLIMQQNSEHFQEILQQFPDVETCQSYK DFQHLKPALRDKISLFTRKEILRARPERILQSLQQVPKQSOEVLLSAGKNSALPGISDGQ QLAKVLLENEVLDLYSQDAETYYTIIVNSFEKEEVLPYREVLKRDLASQLLTSHGHLVQ MERLESALRTRYPGEEGASLWQRRLWXVVENHRLGRHLEGSF5WSLDRSLKTFSRGDKEL PQEFDRIFSMKVGDYSSVFMSPNEGPCYYQCLSHLLYDRPASVDKLFLAKSQLDEELLGS YMERFIEQGVVR

CPn_0435 480908 479475
Phospholipase D superfamily [uncleavable leader peptide]
GVMMSRLRFRLAALGIFFILLVPNSVSAKTIVASDKEKVGVLVVDNSVEAFQOILDCIDH
ANFYVELCPCMTGGRTLKEMVDHLEARMDLVPELCSYIIIOPTFIDAEDQKLLKALKERH
PNRFFYVFTGCPPSTSILAPNVIEMHIKLSIIDGKYCILGGTNFEEFMCTPGDEVPEKVD
NPRLFVSGVRRPLAFRQQIIMLRSTAFGLQLREEYHKQFAMWDYYAHHMWFIDNPEQFAG
ACPPLTLEQAEETVFPGFDKHEDLVLVDSSKIRIVLGGPHDKQPNPVTQEYLKLIQGARS
SVKLAHMYFIPKDELLNALVDUSHNHGVHLSLITNGCHELSPAITGPYAWGNRINYFALL
YGKRYPLWKKWFCEKLKPYERVSIYEFAIWETQLHKKCMIIDDEIFVIGSYNFGKKSDAF
DYESIVVIESPEVAAKANKVFNKDIGLSIPVSHGDIFSWYFHSVHHTLGHLQLTYMPA

481633 480902 CPT_U436 #161633 #016926

[pla-Lipoate Protein Ligase-Like Protein FYVCYMKVRIVDSGKSSAASHMAKDRDLLESLODGELILHLYEWENPCSLTYGHFMRPEK FLLSNYADLGLDAAVRPTGGGFVFHKGDYAFSVLMSATHPSYSSSVLENYHTVNSFVAKV LEKVFRIGGMLAPEDENSSSRDSGNFCMAKTSKYDVLFGDKKIGGAAQRKVQQGFLHQGS LFLSGSSSEFYQRFLKPEVLEEIIEQIQIHAFFPLGLEAADEVLQEARQQVKEAFIKLFC

CPn_0438 485455 484334 ycbF-PF-loop superfamily ATPase NLTLPMPPQVREIMQQTVIVAMSGGVDSSVVAYLFKKFTNYKVIGLFMKNWEEDSEGGLC NEITEMPHOVERIMOQTU VAMSGAVSSVAYLFKKFTNYKVIGLFMKNWEEDSEGGLC SSTKDYEDVERVCLOLDIPYYTVSFAKEYRERVFARFLKEYSLGYTPNPDILCNREIKFD LLOKKVQELGGDYLATGHYCRLNTELQETQLLRGCDPQKDQSYFLSGTPKSALHNVLFPL GEMNKTE/RAITAQAALPTAEKKDSTGICFIGKRPFKEFLEKFLPNKTONVIDMDTKEIV GOHQGAHTYTTIGQRRGLDLGGSEKPCYVVGKNIEENSIYIVRGEDHPQLYLRELTARELN WFTPPKJCGCHCSAKVRYRSPDEACTIDYSSGDEVKVRFSQPVKAVTPGQTIAFYQGDTCL

CPn_0439 485523 486077
No robust homolog present in Genebank/EMBL as of 11/7/98
IISSNNP/LFVSSTLMGVFPSSLPEESADLFITNKEIVALGEKGNVFLTHSIPMHIAAIT
ILVIVALAGIAIICLGCYSQSILLIAVGIVLTILTTLLCLQALVGFIKFIRQLPQQLHTTV OF IREK IRPESSLOLVTNAORKTTODTLKLYEELCDLSOKEFKLOSTLYOKRFELSHKNE

CPn_0440 486081 486740
NO FOULTH HOMOLOGIPESENT IN GENEDARK/EMBL as Of 11/7/98
LATERCEMBATSVAPSFVERSCELGHATEVLHIPPNAYITOPHPIPNAPMETERSKLSTKH
TILFTALTELLTEKSTI FAGYAGYTGMVICOTOLOTIVLTILLALLLATPLKNKOTGTKL
IDDIOCSISSIGDGFVORYGLMFOTTEKVHLIFELTTYNDEFTPILNETEAKKESIQNLEL
KITEK GEFLAGKOPKRKOROKSPMPOTKHLUERNPYTLFDC

486375 487838

GSGVIDVPMIPSEG

EKUM LEGYTINETGEHÖEKAWELMAGAKWILYOLULUK LYEMBE IDAKALEACURGEA AMELIĞUMEKPEHLIYAYEMINIMEN LICACIADEWILEWIACINGAN PELDEWOYN AMELIĞUMEKPEHLIYAYEMINIMEN LICACIADEWILEWIACINGAN PELDEWOYN CEUTORY TAMENTER BELLEMENTER LEGALADEWILEWIACINGAN PELDEWOYN CEUTORY TAMENTER BELLEMENTER BELLEMENTER BELLEMENTER BELLEMENTER CANADAM CEUTORY TAMENTER BELLEMENTER BELLEMENT

RUTRERKKLHKNHLIGGPGIFEY/REIDANVKUTEWEGJFIKGFYGWGIGNDISIADDH NNNKTSHTFKTSAFFOGSAVMNF

489528 487992

ttel_u442 CT006 hypothetical protein NILPOTDMCFKNICKOCSQLYLMXIFPERILARKLKNCAKSYPRTALTIEVLVSSVLGAL KVILIPCASTYAALTLPLRALFNAIKTKSCQHLASYAMAWLLHILTIAVIIGLVFSLVFI PPPVVFTSLGLLMSVTT'SVTLFG'/HKNLFPPYEPPPSRPHTPPPFADEYVPLISESYFD

12000

CTOUS hypothetical protein
USMSOPPINPLOOPOVPAAASPSOOPSVVKRLKTSSTGLFKRFITIPDKYPKMRYVYDT
GIIALAAIAILSILLTASGNSLML/ALAPALALGALGVTLLISDILDSPKAKKIGEAITA
IVVPIIVLAIAAGLIAGAFVASSTMLVFANPMFVMGLITVGLYFMSLNKLTLDYFRREH
LLRMEKKYGETAEPILTYPSADDAKKIAVEKKKDLSASARWEEHEASORQDARRRIGRE
AQGSFFYSSRNPEHRRSFGSLSRFKTKPSDAASTRPASISPFFKDFOPPHFKDLRSSSF GSGASSAFTPIMPASSRSPNFSTGTVLHPEPVYPKGGKEPSIPRVSSSSRRSPRDRQDKQ QQQQNQDEEQKQQSKKKSGKSNQSLKTPPPDGKSTANLSPSNPFSDGYDEREKRKHRKNK

CPI_0444 490266 494507

pmp_6-POlymorphic Outer Membrane Protein

KAFPQRHMKYSLPWLLTSSALVFSLHPLMAANTDLSSSDNYENGSSGSAAFTAKETSDAS
GTTYTLTSDVSITNVSAITPADKSCFTNTGGALSFVGADHSLVLQTIALTHDGAAINNTN

TALSFSGFSSLLIDSAPATGTSGGKGAICVTNTEGGTATFTDNASVTLQKNTSEKDGAAV

SAYSIDLAKTTTAALLDQNTSTKXGGALCSTANTTVQGNSGTVTFSSNTATDKGGGIYSK

EKDSTLDANTGVVTFKSNTAKTGGAWSSDDNLALTGNTQVLFQENKTTGSAAQANPEGC
GGAICCYLATATDKYGLAISONQCMSFTSNTTATNAGGAIYATKCTLDGNTTLTFDONTAT
ACCGGAIYTETEDFSLKGSTGTVTFSTNTAKTGGALVSKGNSSLTGNTNLLFSGNKATGP
SNSSANQEGCGGAILAFIDSGSVSDKTGLSIANNQEVSLTSNAATVSGGAIYAKTCTLTG
NGSLTFDGNTAGTSGGAIYTETEDFTLTGSTGTVTFSTNTAKTGGALYSKGNNSLSGNTN

LLFSGNKATGPSNSSANQEGCGGAILSFLESASVSTKKGLWIEDNENVSLGSNTATVSGG
AIYATKCALHGNTTLTFDGNTAETAGGAIYTETEDFTLTGSTGTVTFSTNTAKTAGALHT
KONTSFTKNKALVFSGNSATATATTTTDQECCGGAILCAISESDIATKSLTLTENESLSF
INNTAKRSGGGIYAPKCVISGSESINFDONTAETSGGAIYSKNLSITNAGFVSFTNNSGG
KGGAIYIADSGELSLEAIDGDITFSGNRATEGTSTFNSIHLGAGAKITKLAAPGHTIYF
YDPITMEAPASGGTIEELVINPVVKAIVPPPQPKXGPIASVPVVPVAPANPNTGTIVFSS
GKLPSQDASIPANTTIILNQKINLAGGNVVLKEGATLQVVSFTQQPDSTVPMDAGTTLET
TTNNTDGS IDLKNLSVNLDALDGKRMITIAVNSTSGGLKISGDLKFRNNGSGFVDNPCL
KANLNLPPLDLSSTSGTVNLDDFNPIPSSMAAPDYGYQGSWTLVPKVGAGGKVTLVAEWQ TTTNNTDGSIDLKNLSVNLDALDGKRMITIAVNSTSGGLKISGDLKFHNNEGSFYDNEGL KANLNLPFLDLSSTSGTVNLDDFNPIPSSMAAPDYGYQGSWTLVPKVGAGGKVTLVAEWQ ALGYTPKPELRATLVPNSLWNAYVNIHSIQQEIATAMSDAPSHPGIWIGGIGNAFHQDKQ KENAGFRLISRGYIVGGSMTTPQEYTFAVAFSQLFGKSKDYVVSDIKSQVYAGSLCAQSS YVLELHSSLRRHVLSKVLPELPGETPLVLHGQVSYGRNHHMMTTKLANNTQGKSDWDSHS FAVEVGSLPVDLNYRYLTSYSPYVKLQVVSVNQKGFQEVAADPRIFDASHLVNVSIPMG LITECHESAKPPSALLTLGYAVDAYRDHPHCLTSLTNGTSWSTFATNLSRQAFFAEASGH LKLEGHGLDCFASGSCELRSSSRSYNANCGTRYSF

494739 497579

CPL_0445 494739 497579
pmg_7-Polymorphic Outer Membrane Protein
pmg_7-Polymorphic Outer Protein
pmg_7-Polymorphic Outer Protein
pmg_7-Polymorphic Outer Protein
pmg_7-Polymorphic Outer Membrane Protein
pmg_7-Polymorphic Outer Protein
pmg_7-Polymorphic Outer Protein
pmg_7-Polymorphic LARGGFSVRAANHFQVNPHMEIFGQFAFEVRSSSRNYNTNLGSKFCF

CPA_9446 497602 500415

mp_8-Polymorphic Outer Membrane Protein
LIEBRHLSMKIPLHKLLISSTLVTPILLSIATYGADASLSPTDSFDGAGGSTFTPKSTAD
ANGTNYVLSGNVINDAGKGTALTGCCFTETTGDLIFTGKGYSFSFNTVDAGSNAGAAAS
TTADKALTFTGFSNLSFIAAFGTTVASGKSTLSSAGALNLTDNGTILFSONVSNEANNNG
GAITTKTLSISGNTSSITFTSNSAKKLGGAIYSSAAASISGNTGGLVFMNNKGETGGGAL
GFEASSSITQNSSLFFSGNTATDAAGKGGAIYCEKTGETPTLTISGNKSLTFAENSSVTQ
GGAICAHGLDLSSAGEPTLFSNNRCCNTAAGKGGAIAIADSGSLSLSANGGDITFLGNTLT
STSAFTSTRNAIYLGSSAKITNLRAAQOGSIYFYDPIASNTTGASDVLTINOPDSNSPLD
YSGTIVFSGCKLSADGAKAADNFTSILKQPLALASGTLALKGNVELDVMGPTQTEGSTLL
MQPCTKLKADTEAISLTKLVVDLSALEGNKSVSIETAGANKTITLTSPLVFQDSSGNFYE
SHTINQAFTQPLVVFTAATAASDIYIDALLTSPVQTPEPHYGYQGHWEATWADTSTAKSG
TMTWYTTGYNPNEPERRASVVPDSLWASFTDIRTLQOIMTSQANSIYQGNGLMASGTANFF
HKDKSGTNQAFRHKSYGYIVGGSAEDFSENIFSVAFTQLFGKDKDLFIVENTSHNYLASL
YLQHRAFLGGLPMPSFGSITDMLKDIPLINAQLSYSYTKNDMDTRYTSYPEAQGSWTNN
SGALELGGSLALYLPKEAPFFQVFPFLKFQAVYSRQQNFKESGAEARAFDDGDLVMCSI SGALELGGSLALYLPREAPFFGGYFPFLKFQAVYSRQONFKESGAEARAFDDGDLVNCSI PVGIRLEKISEDEKNNFEISLAYIGDVYRKNPRSRTSLMVSGASWTSLCKNLARQAFLAS AGSHLTLSPHVELSGEAAYELRGSAHIYNVDCGLRYSF

500541 503351

CPD_0447

pm_02+Polymorphic Outer Membrane Protein

PVKPPIAL/MKSSLHWFLISSSLALPLSLNFSAFAAVVEINLGPTNSFSGPGTYTPPAQT

TNAD7TI/NLTGDVSITNAGSPTALTASCFKETTGNLDFGOFGYGPLLQNIDAGANGTFT

TNAD7TI/NLTGDVSITNAGSPTALTASCFKETTGNLDFGOFGYGPLLQNIDAGANGTFT

TNAAD7LLSPSGPGYJELLLQTPTNATTGTGAIKGTGAGSIQQNYSCYFGQNFSNDNGGALQ

GGSIGLGLNPNLTPAKNATOKGGALYGTGGITINITHASAFSENTAANNGGAIYTEAS

SPIGSHAAIGEINNSVTNTSATGAAIYGCSTBAPKFV/LTLDDNGGLNFIGNTAITSGGAI

GTDNLYLGSGPTLFKNNSAIDTAAPLGGAIA I ADGGSLLGSALGGITFFENTAVKGAS

GGGTTFPHSINIGNTNAKIVGLRAAQGANTIYTYDPITTGITAALDAINLNGPDLAGNPA

GGGTTFPHSINIGNTNAKIVGLRAAQGANTIYTYDPITTGITAALDAINLNGPDLAGNPA

MAGTTLFFARXITINNAKIVGLRAAGGANTIYTYDPITTGITAALDAINLNGPDLAGNPA

HAGGTTLFTAXGTTANDPANIHIIDLAADPLEKNFIHMG/G/GNWALWQEDFATKSKAA

TUTMPTTYYNNNHERPITTLVANTHEGTTVONFIHMFYNGPGFTEGINGFEGINFF

HKDTPKITTYNNHHERPITTLVANTHEGTTVONFIHMFYNGPGFTEGINFEGINFF

HKDTPKITTYNNHHERPITTLVANTHEGTTVONFIHMFYNGPGFTEGINFEGINFF

HKDTPKITTYNNHHERPITTLVANTHEGTTVONFIHMFYNGPGFTEGINFEF

HKDTPKITTYNNHHERPITTLVANTHEGTTVONFIHMFYNGPGFTEGINFEF

HKDTPKITTYNNHHERPITTLVANTHEGTTVONFIHMFYNGPGFTEGINFEF

HKDTPKITTYNNHHERPITTLVANTHEGTTVONFIHMFYNGPGFTEGINFEF

HKDTPKITTYNNHHERPITTLVANTHEGTTVONFIHMFYNGPGFTEGINFER

HKDHLATLSSIBGLARLAFFEFEKNARVHGOFFFKFFNTTHINSBEGGGGWNDGGC

AUELAGGLGFTALSBEGLARLAFFFEKNARVYHGFTEFFTTHINSBEGGGGGWNDGGC ALELACCLEPPALCHEDILFIIAY FOF LKVDACY LHODGEKFPNTTTLVRSEDSGDL INVSVP TO LTFERFORNERACY LATV LYV NOVYRKNER TITALL I INTOWKT TYTINGROAG IGRA

GIFYAFSPNLEVTCHLSMEIRGSSRCYNADLAGKFOF

CPn_0448 504876 503698

*yx7G_Bs_2 Hypothetical Protein
FiQPSRREIHEWKCILLGSSLRMEMMSFFQQPEQCHFDVVGSFLRPESLTRARSDFEEGR
FIVYEQMRVVEDAAIRNLIKKOTEAGLIFFTDGEFRRYSWDFDFMWGFHGVDRRRDSNDPE
IGVYLKDKISVSKHPFIEHFEFVKTFEKGNAKAKQTIPSPSQFFHEMIFAPNLKNTRKFY
PTNOELIDDIVYYFQVIQDLYAAGGRRUDLDDCAWGRLLDIRAPSWYGVDSHDRLQEIL
. WYINNIAMWS DELGGERREINDELGGERRAFSWYGVDSHDRLQEIL
. WYINNIAMS DELGGERREINDELGGERRAFSKYGVDSHDRLQEIL CGFASCEGDHRMTEEEQWKY IAFVKE IAKLIWG

CPn_0449 507231 505330

pmp_10-PMP_10 (Frame-shift with 0451)

EAYTGFRGGGGISFSNNIVQGTTAGNGGAISILAAGECSLSAEAGDITFNGNAIVATTPQ

TTKRNSIDIGSTAKITNLRAISGHSIFFYDPITANTAADSTDTLNLNKADAGNSTDYSGS

IVFSGEKLSEDEAK/ADNLTSTLKQPVTLTAGNLVLKRGVTLDTKGFTQTAGSSVIMDAG

TTLKASTEEVTLTGLSIPVDSLGEGKKVVIAASAASKNVALSGPILLLDNQGNAYENHDL

GKTQDFSFVQLSALGTATTTDVPAVPTVATPTHYGYQGTWGMTWVDDTASTPKTKTATLA

WNNTGYLENPERGGPLVPNSLWGSFSDIQAIQGVIERSALTLCSDRGFWAAGVANFLDKD

KKGEKRKYRHKSGGYAIGGAAQTCSENLISFAFCQLFGSDKDFLVAKNHTDTYAGAFYIQ

HITECSGFIGCLLDKLPGSWSHKPLVLEGQLAYSHVSNDLKTKYTAYPEVKGSWGNNAFN

MMLGASSHSYPEYLHCFDTYAPYIKLNLTYIRQDSFSEKGTEGRSFDDSNLFNLSLPIGV

KFEKFSDCNDFSYDLTLSYVPDLIRNDPKCTTALVISGASWETYANNLARQALQVRAGSH

YAFSPMFEVLGOFVFEVRGSSRIYNVDLGGKFQF

508121

CPI_0430
pmp_10-Polymorphic Outer Membrane Protein
SGFMKSQFSWLVLSSTLACFTSCSTVFAATAENIGESDSFDGSTNTGTYTPKNTTTGIDY
TLTGDITLQNLGDSAALTKGCFSDTTESLSFAGKGYSLSFLNIKSSAEGAALSVTTDKNL SLTGFSSLTFLAAPSSVITTPSGKGAVKCGGDLTFDNNGTILFKQDYCEENGGAISTKNL SLKNSTGSISFEGNKSSATGKKGGAICATGTVDITNNTAPTLFSNNIAEAAGGAINSTGN CTITGNTSLVFSENSVTATAGNGGALSGDADVTISGNQSVTFSGNQAVANGGAIYAKKLT LASGGGGVSPFLTI

CPn_0451 508158 511058
pmp_10-PMP_10 (Frame-shift with 0451)
KTQRVKIKILDSCFVIFNLIYLFCFYIDANSSLKNKSITMKTSIPWVLVSSVLAFSCHLQ

KTQRVKIKILDSCFVIFNLIYLFCFYIDANSSLKNKSITMKTSIPWVLVSSVLAFSCHLQ
SLANEELLSPDDSFNCNIDSGTFTPKTSATTYSLTGOVFFYEPGKGTPLSDSCFKQTTDN
LTFLGKNFLSLFFGFDAGTHAGAAASTTANKNLTFSGFSLLSFDSSPSTTVTTGGGTLSS
AGGVNLENIRKLVVAGNFSTADGGAIKGASFLLTGTSGDALFSNNSSSTKGGAIATTAGA
RIANNTGYVRFLSNIASTSGGAIDDEGTSILSNNKFLYFEDNAAKTTGGAICNTKASGSP
ELIISNNKTLIFASNVAETSGGAIHAKKLALSSGGFTEFLRNNVSSATPKGGAISIDASG
ELSLSAETGNITFVRNTLITTGSTDTPKRNAINIGSNGKFTELRAAKNHTIFFYDPITSE
GTSSDVLKINNGSAGALNPVGGTILFSGGTLTADELKVADNLKSSFTOPVSLSGGKLLLQ
KGVTLESTSFSQEAGSLLGMDSGTTLSTTAGSITITNLGINVDSLGLKQPVSLTAKGASN
KVIVSGKLNLIDIEGNIYESHMFSHDQLFSLLKITVDADVDTNVDISSLIPVPAEDPNSE
YGFQCOWNVMYTDTATNTKEATATWTKTGFVPSPERKSALVCNTLWGYFTDIRSLQQLV
EIGATGMEHKQGFWSSMTNFLHKTGGENRKGFRHTSGGYVIGGSAHTPKDDLFTFAFC
LFARDKDCFIANNNSRTYGGTLFFKHSHTLOPONYLRLGRAKFSESAIEKFPREIPLALD LFARDKDCF IAHNNSRTYGGTLFFKHSHTLØPONYLRLGRAKFSESAIEKFPREIFLALD
VOVSFSHSDNRMETHYTSLPESEGSWSNECIAGGIGLDLPFVLSNPHPLFKTFIPOMKVE
WYYYSQNSFFESSSORGFSIGRLLLNLSIPVGKFVGGDIGDSYTYDLSGFFVSDVYRNN
PQSTATLVMSPDSWKIRGGNLSRQAFLLRGSNNYVYNSNCELFGHYAMELRGSSRNYNVD

CPn_0452 511304 512860
pmp_12-Polymorphic Outer Membrane Protein (truncated)
FNEETMTILRNFLTCSALFLALPAAAQVVYLHESDGYNGAINNKSLEPKITCYPEGTSYI

FNEETMTILRNFLTCSALFLALPAAAQVYLHESDOYNGAINNKSLEPKITCYPEGTSYI
FLDDVRISNVKHOGEDAGVFINRSSNLFFMGRROHTFHNLMTEGFGAAISNRVGDTTLT
LSNFSYLAFTSAPLLPQGQAIYSLGSVMIENSEEVTFCGNYSSWSGAAIYTPYLLGSKA
SRPSVNLSGNRTLVFRENVSGQYGGAISTHNLTLTTTRGPSCFENNHAYHDVNSNCGAIAI
APGGSISISVKSGDLIFKGNTASQOGNTIHNSIHLQSGAQFKNLRAVSESGVYFYPDISH
SESHKITDLVINAPECKETVEGTISFSGLCLDDHEVCAENLTSTILQDVTLAGGTLSLSD
GVTLQLHSFKQEASSTLTMSPGTTLLCSGDARVQNLHILIEDTDNFVPVRIRAEDKDALV
SLEKLKVAFEAYWSVYDFPOFKEAFTIPLLELLGPSFDSLLLGETTLERTQVTTENDAVR
GFWSLSWEEYPPSLDKDRRITPTKKTVFLTWNPEITSTP

513156 516152

CPn_0453 513156 516152
pmp_13 -Polymorphic Outer Membrane Protein
NCVLLYLFFYSLSLICRIIWHLLYVQMKTSIRKFLISTTLAPCFASTAFTVEVIMPSENF
DCSSGKIFPYTTLSDPROTLCIFSGDLYIANLDNAISRTSSSCFSNRAGALQILGKOGVF
SFLNIRSSADGAAISSVITQNPELCPLSFSGFSQMIFDNCESLTSDTSASNVIPHASAIY
ATTPMLFTNNDSILFQYNRSAGFGAAIRGTSITIENTKKSLLFNXNGSISNCGALTGSAA
INLINNSAPVIFSTNATGIVGGAIYLTGGSMLTSGNLSGVLFVNNSSRSGGAIYANGNVT
FSNNSDLTFQNNTASPONSLPAPTPPPTPPAVTPLLGYGGAIFCTPPATPPPTGVSLTIS
GENSVTFLENIASGJGGALYGKKISIDSNKSTIFLGNTACKGGAIA IFESGELSLSANQC
DILFNKNLSITSGTPTRNSIHFGKDAKFATLGATQGYTLYFYDPITSDDLSAASAAATVV
VNPKASADCAYSGTIVFSGETLTATEAATPANATSTLNQKLELBOGTLALRNGATLNVHN
FTQDERSVVIMDAJTTLATTNGANNTDGAITLINKLVINLDSLDGTKAAVVNVOSTNGALT
ISGTLGLVKNSQCCDNHGMFNKDLQQVPILELKATSNTVTTTDFSLCTNGVQQSPYGYQ
GTWEFFIDTTTHTTGNWKKTGYLPHPERLAPLIPNSLWANVIDLRAVSQASAADGEDVP
GKQLSITGITNFFFANNTGDARSYRHMGGGYLINTYTRITPDAALSLGFGQLFTKSKDYL
VCHGHSNVYFATVYSNITKSLFGSSRFFSGGTSRVTYSRSNEKVKTSYTKLPKGRCSWSN
NCWLGELEGNLPTLASRILNIKQIIPPVKAREVAYATHGGIGENTPEGRIFGHGHLINVA
VPVGVPFGKNSHNRSDFYTIVAYAPDDVRHNPDCDTTLPINGATWTSIGNNLTRSTLLV VPVGVPFGKNSHNRFDFYTTIVAYAPDVYRHNPDCDTTLPINGATWTSIGNNLTRSTLLV QASSHTSVNDVLEIFGHCGCDIRRTSRQYTLDIGSKLRF

CPI_9454 516179 519115
pmp_14-Polymorphic outer Membrane Protein
GMPLOFKGSFOCLIALICSASCAFACTRICGIEVPPTTNOGEETLLTSDEVCHNELGASF
SGSFINISSINISGLIALICSASCAFACTRICGIEVPPTTNOGEETLLTSDEVCHNELGASF
SGSFINISSINISGLIALICSASCAFACTRICGIEVPPTTNOGEETLLTSDEVCHNOCTGL
GGLIVGKDIVFOSIKOLIFTTNRVAYSPASVITISATPATTTVTTGASALQPTDBIJTVENI
GGLIVGKDIVFOSIKOLIFTTNRVAYSPASVITISATPATTTVTTGASALQPTDBIJTVENI
GGLIVGKDIVFOSIKOLIFTTNRVAYSPASVITISATPATTTTGASALQPTDBIJTVENI
GGLIVTANISCANISCASTATALABORICALABORICALIFOTELEKNNOCKCHTSVINISTENDAG
ALVARTONIVANONALLILDINTAARNOCIALICARVILNIQGEGFIEFSRNRARKKORIFICHPY
DPITHICHTTGESNKOLITINAN ANGRIVVFTCVGLISTELLLPANTTFTLLATIVK LAGGE
LKITIMIAVVNIGENIVGA GAGUTISKKYTLGGATPTGAPAAVDETIGKATDFTSELKFD
PVGAZNANTYNINTGALALDBIHAVTBIJTMINTGASTAFITIALGENTVIKKRIGHEGGE
EKITIMIAVVNIATPNINTGALALDBIHAVTBIJTMINTGASTAFITIALGENTVIKKRIGHEGGE
LKITIMIAVVNIATPNINTGALALDBIHAVTBIJTMINTGASTAFITIALGENTVIKKRIGHEGGE
LKITIMIAVVNIATPNINTGALALDBIHAVTBIJTMINTGASTAFITIALTRIANTVIKKRIGHEGGE
LKITIMIAVVNIATPNINTGALALDBIHAVTBIJTMINTGASTAFITIALTRIANTVIKKRIGHEGGE
LKITIMIAVVNIATPNINTGALALDBIHAVTBIJTMINTGASTAFITIALTRIANTVIKKRIGHEGGE
LKITIMIAVVNIATPNINTGALALDBIHAVTBIJTMINTGASTAFITIALTRIANTVIKKRIGHEGGE
LKITIMIAVVNIATPNINTGAGALABORITANISTER PROTEINTERGAGARVERTTPPLGHEGGE
LKITIMIAVVNIATPNINTGAGALABORITANISTERGAGARVERTTPPLGHEGGE
LKITIMIAVVNIATRIANTGAGARVERTTPPLGHEGGEGGG

YGGYOAALI:MMYTDHTTLGLDFGQLYGKTNANPYDSRCSEQMYLLSFFGQFPIVTQKSEA LI:BWKAAYGYGKNHLNTTYLRPDKAPKSGGOWHNNSYYVLISAEHFFLWCLLTRPLAQA WDLSGFISAEFLGGWGKFTETGDLQRSFSRGKGYNVSLPIGGSSOWFTPFKKAPSTLTI KLAYKPDIYRVNPHNIVTVVSNQESTSISGANLRRHGLFVQIHDVVDLTEDTQAFLNYTF

CPn_0455 520363 519458
No robust homolog present in Genebank/EMBL as of 11/7/98
***PRESENTATION OF PRESENTATION OF THE PRINCIPLE FOR VIOLETATION OF THE PRINCIPLE FOR VIOLETATION OF THE PERSONNAL OF THE PRINCIPLE FOR VIOLETATION OF THE PERSONNAL ASGCKVVAFDDEHLPWVSSHIAYAEEIREKQEQTMQGSLTEEQLGALLCNTVSTEKNLAF ALDAVIKQSVWRFRNPDLFAYEREALEASVTDALVSYVSNLDMIPYTSSQGIVIEDSSIV RTSQEHTLIVNCAAFDKLASQIEFLCPSDVLPISGKDPLISDDEDEELNPKVSSAADSKD

CPn_0456 521568 520327

No robust homolog present in Genebank/EMBL as of 11/7/98
IPCTFESKRKFLMTHCLHGWFSVVRHHFVQAFNFSRPLYSRITHFALGVIKAIPIVGHLV
MCVDMLISHCFERGVSHPGFPSDIAPILKVEKIAGDOHISRIEMOLKSLRKTIEVEDLDK
VHGQYQENPYADMASSSVLKLDKGVHVSELGKAFSRVRNRITRSYSYAPTPQLDSIAIVG
IDLVSPEEQENLVRLANEVIGLYPKSKTTLYLLIDFNKEWVGDISSDKEKQLRSLGLHSE
VQCLSVLEPQABEGEDTKHFDLMVGCYGKDSVLREGKILQOALGTSLGTVPWVMVHTLP
SRYRSRLSLPINTEKDKTELYKEISRTHHQLHTLGMGJGAQDSGLLLDRQRLHAPLSQGS
HCHSYLADLTHEELKILLFSAFVDAKNISKKELREVSLNFANDTSVECGCAFYF

523886 522120 CPn 0457

CPn_0457 \$23886 \$522120

No robust homolog present in Genebank/EMBL as of 11/7/98
VFLPSRVMASCLSAWFSIVREHFYRAFDFSLPFCARITEFVLGVIKGIPVVGHIIVGIEW
LVSRYLESFVTKEFFVSDVVSLLKTEKVAGRDHIARVVETLKRQRVAVAPEDEDKVHGKI
EVHPFGGIOPVEVLTLYPEVODATIGLAFSKIRNTVRQAYLQAPPRKLGKIYIIGNDMNP
FEVDDFLHLARLCNETQRLYPDATISLYLTASGGRNAMDKKNRKLLSDCELNPKIACLDF
NQGDVVKQATCDCWMYYHGENDQGTLNQIQELEKSGEFTWIHVQQKFLSQSLWDFSPF
SSLEMKGDKEKALEYSELEKEQLYSRLVYVGERSSVLSLGFGDSRSGILMDFRRYHAPLS
EGHYCHSYLADLENPGLQKTLIAAFLNPKELSSTILOPISLNLILNSKTYLRQHFGFFER MSRSDRNVVVVVCSMWGTDWKEEPSFOHFIMELECRGYSHFNIFAFRSNSMCVEERRIL NESSQEKAFTMIFCEDSVSQGDIRCLHLASEGMLCGKECYAVDVYTSGCANFMMEEVLTL ERESNLWNRKHGLWKREVRKQKQEAALDQDESEIYVCNQLTAQQNFACS

CPn_0458 526344 524236
No robust homolog present in Genebank/EMBL as of 11/7/98
YEBECYLKLFVSNFIFFVVMPIPYISSWISTVRQHFVKAFDFSRPFCSRVTNFALGVIKA YEECCYLKLFVSNFIFFVMPIPYISSWISTVROHFVKAFDFSRPFCSRVTNFALGVIKA
IPHVGHIVNGMENLVSSCVAGIITRSSFTSDVVQIVKTEKALGRDHISRVAEILQRERGT
IPHNQGKVHGKFFVCPFGRLKSEETLKLKPGEREGTLDTVFSPIRTRYTRAYLQAPRPE
IRTISIVGSKLKTPQDFSQFVSLANETQRLHPEALVCLYLTGLNRESQMCDTTTAEKKQY
LHNSGLDSRIQCKDSKEDDAGSPENPELWIGTYSREQOHNIDGQYIQCLGKSADFIPWI
HVTEDTKDFYYPFNFTSYSHTRQSTDPTSFPRLPESEGDKDSLYGQLSRSYHHEYWIGLG
LKPEDAGLLMDPDRIYAPLSQGHYCHSYLADIENEDLRTLVLSPFLDPGNLSSEDLRPVA
FNIARLPLELDSLFFRLVAGQQESRNIVTLAHGTPRPEDLDPDSNNILTRRLGMSGYSYL
NEPSYKSRKMIVKERGFFGDRSEGKSFTLILFEDPISAADFRCLQLAARGMVAKDLPSVA
DECASGCSCIQFSEMQSPQAIEYRGWEARVEDEAGERAREPVIYSQDQLSSMLTTQQNFV
FSLDAVVKQAIWRFRSKGLLTMERKALGEEFLTAIFSYLGSQERNENMGKRTTEEHEVVI
SEEELDRMVQVLPAEVPADSGNDPTRPVPNPDSNPDSSQNEGS

CET10459 527062 526619
NÖ TODUST homolog present in Genebank/EMBL as of 11/7/98
STEKIOMHFOLRNWRTSTINKLREEGSVSFREYFRAYMCDKIVAQKNFLFTLDAVIKQACKR
SQEKLINLFYVESQALGREIKVSLEEYIQSMVGILGSQRTKKSFKFSVDFTPLEQALQERC SSOODEDATATSTATGATASPTDMHEDE

CETE-0460 527840 526992
No. robust homolog present in Genebank/EMBL as of 11/7/98
VIGHLNFALEETPSISVOYQEQEKLSPCDHSPEIGKKKRNNKLESFSTYCSLFMSVKDH
YKLNIG IONSLSGWLLDPYRVCAPLSSPYSCPSYLLDLONKELRSSLLSTFLDPKNLTSE
TFKSVSINFGNSSFGORWSEFLSRVLHDEKEKHVAVVCNDAKLLEEGLSPEALSLLEEDL RESGYSYLNILSVSPEGVSKVQERQILRRDLQGRSFTVMITDLPLGSEDIRSLQLASDRI LVSSSLDAADACASGCKVLVYENPNASWAQELENFYKQVERRR

LVSSIDABLICATION

CPITO461 528647 527844

No robust homolog present in Genebank/EMBL as of 11/7/98
ISIVACPSISSWFTVVRQHFVNAFPFTHEVCSRITNFALGIIKAIPVLGHIVMGIEMLIS
WIPRRTVRHGMFTSDVSSAIRVEQTRGHNCLAFLEAVLSSLRVPISQEDLGKVHGRTPED
FYDJITPFIIVQLLIPDEELSTVDERALTGYAVRSVEKPMIQDLALVGFGLRDSAD
LINFVRLANGVQNHYPHTKVKLYLAKNLADVWDCEISEEEKGQLRALGLDPKIESISLTS

CPn_0462 531124 529037

No robust homolog present in Genebank/EMBL as of 11/7/98
LIFYLFLNLY1ACVRFHFOCWFDPMACY1SIWISTVKOHFIRAFDFTRELGSITNFALG
VIKAIPILGGVVIGOSMLVSTGSARFFGKPAFTSDVASIVIEKTRGYMPLAWYEQYLRQ
LRVRLPEGDLGKIHGKVSRDYVCDRTPQENLNMVPHQYLGELGRAFYGIRNRVTKAYQRV
TPLEVPCLTLVGFDILDPEDDVNFVRLANGIGTQVYPOTOIKLYLISIGNIMOCDCTISQ
EKEQQLRSLGLDAKIKCVSAPALLLQKYLQSENLPSCDLLINYYCKQOSVRDVDSIKSLL
NLSGEHIPAISVTYRPDDPFYSYYFFGSQGTAPDQRIPMSEQEHLQTYTTLSNPRCDR
YAVHLGMEDPFASGVFLDPLRVSAPLSGEVSCPSYLLDLKSEELRCFLLSAFIDPNNSGQG
NPRPMSINFGNSPLGQRWSEFLSRVLHDETEKHVAVVCNNPOLIKKSFPSHSLSLLENEL
EBGGTSYLNIVSVSQERTCVKERRILSSDPSGRSFTVILTDLPEGSSDIRNLQLASDRIL
VGSALDAADACASECKILEYEDPEQEWAQQYASFYPNIDRAGDLQROGIPGEPLGVSAST
RVVLEKDIVPNLNAVIQQAMWKFKKRDLFAVESQALGDDMRRALEGYIGSSLLVEGTIQP
QVACNVNVSFATLDEAVCACDSAQDAPSEENNTDD

532490 531191

CPn_0463 532490 531191

To robust homolog present in Genebank/EMBL as of 11/7/98

EGSPYEKTEQLECTPNCRTPRVN LSTVG LPIDETSHAFVDSMMKG3VGQDAKELYTFLSR

CNEHYQLCLWFGLEGELGFLFDCKMLCAPLSCDHYCHGYLVDLVVQHLKDL LLSMFLDPQ

MIGAGELLKVG INVOIGHSFLAQKOFLIAMVLROETSERVVVVYFKGVLGLPATQVCKLVEE

LNGKDYGYLNLFGGHGDSSFQLLFRKELEGTSGRYFTVICALYLGDTDMRSLQLASERIM

VOREFDLVDAYAARGKILK IDHTMWRFGTFSGRHADFADAVDVGAGFNSREFKLITQANQG

LLEGGELGEGERTFWRGGPLAFCDRVTVTRHFIFMLDAAIKQAVWTHKHGGLIDKECEALD

LYTYGLEGIVGYLEYVTNGHEKTGKGPFIOKEIHADGGPLKEALFPGSDEDVPSTSEDPS

DARLFGLLEFD

CPn_0464 5/2023 5323hh No robust nomolog present in Genebank/EMBL in at 11 7/99 SLETRGRETEICLQLLFFDIGSLKFLQLFSECTALNLER IFAPLRNRVTTEYSRARQPDL HRIAIVYIGVLDSESKILERLISYMSCIYSESOMYLRFFMGKNVNQSAVLSKLHVENLH IRCGFFSEDAVPESEPFDLSIYVHTDRSCPLPTKKRSSSWELQTVELPESIYPQSEFLLM

CPn_0465 533278 532871

TO PER TRANSPORT TO THE TRANSPORT OF THE TRANSPORT RDEGKVHGDLPSAPFF

CPN_0466 533718 536537

pmp_15-Polymorphic Outer Membrane Protein
TSMRFFCFGMLLPFTFVLANEGLQLPLETYITLSPEYQAAPQVGFTHNONQDLAIVGNHN
DFILDYKYYRSNGGALTCKNLLISENIGNVFFEKNVCPNSGGATYAAQNCTISKNONYAF
TTNLVSDNFTATAGSLLGGALFAINCSITNNLGGGTFVONLALINKGGALYTETNLSIKDN
KGPIIIKONRALINSDSLOGGIYSGNSLNIEGNSGAIQITSNSSGSGGGIFSTQTLTISSN
KKLIEISENSAFANNYGSNFNPGGGGLTTFFCTILNNREGVLFNNNQSQSNGGAIHAKSI
IIKENGPVYFLNNTATRGGALLNLSAGSGNSFILSADNGDIIFNNNTASKHALNPPYNN
AIHSTPNMNLQIGARPGYRVLFYDPIEHELPSSFPILFNFETGHTGTVLFSGEHVHQNFT
DEMNFFSYLRNTSELRGGVLAVEGDAGLAVFFFORGGTLLGGGAVITTAGTIPTPSST
PTTVGSTITLNHIAIDLPSILSFQAQAPKIWIYPTKTGSTYTEDSNPTITISGTLTLRNS
NNEDPYDSLDLSHSLEKVPLLYIVDVAAQKINSSQLDLSTLNSGEHYGYQGIWSTYWVET
TIITNPTSLLGANTKHKLLYANWSPLGYRPHPERRGEFITNALWOSAYTALAGLHSLSSW
DEKGHAASLQGIGLLVHQKDKNGFKGFRSHMTGYSATTEATSSQSPNFSLGFAGFFSKA
KEHESONSTSSHHYFSGMCIENTLFKEBITRLSVLAYMFTSEHTHTMYGQLLEGNSGGSF
HNHTLAGALSCVFLPQPHGESLQIYPFITALAIRGNLAAFQESGDHAREFSLHRPLTDVS
LPVGIRASMONHRVPLWALTEISYRSTLYRQDPELHSKLLISQGTWTTQATPVTYNALG
IKVKNTMQVFPKVTLSLDYSADISSSTLSHYLNVASRMFF

536528 539434

CPH_0467 536528 539434

pmp_16-Polymorphic Outer Membrane Protein

NeiltisDonkrikEpluskTPPkflfylghfTacMfgMTPAVYSLQTDSLEKFALERDE

EFRTSFPLLDSLSTLTGFSPITTFVGNRHNSSQDIVLSNYKSIDNILLLWTSAGGAVSCN

NFLLSNVEDHAFFSKNLAIGTGGAIACGAGCTITKNRGPLIFFSNRGLNNASTGGETRGG

AIACNGDFTISONQGTFYFVNNSVNNWGGALSTNGHCRIQSNRAPLLFFNNTAPSGGGAL

RSENTTISDNTRPIYFKNNCGNNGGAIGTSVTVAIKNNSGSVIFNNTTALSGSINSGNGS

GGAIYTTHISIDDNTRDIILFNNNVIRGGGAITL

LQDSTCLLFAEGGNIAFQNNEVFLTTFGRYNAIHCTPNSNLQLGANKGYTTAFFDPIEHQ LQDSTCLLFAEQONIAFQNNEVFLTTFGRYNAIHCTPNSNLQLGANKGYTTAFFDPIEHQ
HPTTNPLIFNPNANHQSTILFSSAYIPEASDYENNFISSSKNTSELRNGVLSIEDRAGWQ
FYKFTQKGGILKLGHAASIATTANSETPSTSVGSQVIINNLAINLPSILAKGKAPTLWIR
PLQSSAPFTEDNNPTITLISGPLTLLINEENRDPYDSIDLSEPLQNIHLLSLSDVTARHINT
DNFHPESLNATEHYGYQGIWSPYWVETITTTNNASIETANTLYRALYAWMTPLGYKWPE
YQGDLATTPLWQSFHTMFSLLRSYNRTGDSDIERPFLEIQGIADGLFVHQNSIPGAPGFR
IQSTGYSLQASSETSLHQKISLGFAQFFTRTKEIGSSNNVSAHNTVSSLYVELPWFQEAF
ATSTVLAVGYGDHHLHSLHPSHQEDQAEGTCYSHTLARAIGCSFPWQKSYLHLSPFYQAI
AIRSHQTAFEEIGDNPRKFVSQKFFYNLTLPLGIQGKWQSKFHVPTEWTLELSYQPVLYQ
QNPQIGVTLLASGGSWDILGHNYVRNALGYKVHNQTALFRSLDLFLDYQGSVSSSTSTHH
LOAGSTLKF LQAGSTLKF

CPn_0468 539608 540432
pmp_17-Polymorphic Outer Membrane Protein
IYKLLDNKLMIFYDKLYFHIKVWMFMRPICLSILSTALCCSLSGNEVPNLASCQMSRKDI
SAFHTSPSFRLNVTPEPLVSSFRPSNLLNGFGHDITODITITGNSINSVIDNYHYEDGG
ILACKNLFISENKGNLSFERNSSHSSGGALYSVRECWISKNQNYSFISNAASLATTTTSG
FGGAIHALDSYITNNLGEGQFLDNVSKNRGGAIYVGVSLSITDNLGPIVIKKNQTLEDSS
FGGGIFCRAVNIERNYQNIQINDNASGQGVVYFLP

0469 540399 541460 17-Polymorphic Outer Membrane Protein (Frame-shift with

0469)
(CFRTRGGIFSALGVIISSNKEIIEISNHSASSINTASGKLYPGGGGIMCTSLVIENNPKG
LIFNNKTAALSGGAIHTRSFIFQNNGPTAFINNSATSGGALINLSGIGSTPQNFFLSADY
GDILFNNNTITSSSPQPGYNALYAAPGINLKLGARQGYKILFYDPIDHDQTTTDPIVFN
YEPHHLGTVLFSGINVDSNATNPLNFLSKFSNSKLERGVLAIEDRAAISKTLSQTQGI
LRLGNAALIRTKGPGSSINFNAIAINLPSILQSEASAPKFWIYPTLTGSTYSEDTSSTIT LSGPLTFLNDENENPYDSLDLSEPRKDIPPPLPPRCDCKKNRYFESHCRSHELR

541357 542532

 $_{\rm CLC}$ $_{\rm CC}$ $_{\rm$

USLNLERISPLLYLLDVTAKKIDTSNLIVEAMNLDEHYGYQGIWSPYWMETTTTTSSTVP EQTNTNHRQLYYDWTPVGYRPNPERHGEFIANTLWGSAYNALLGIRILPPQNLKEHDLEA SLQGLGLINQNHREGRKGFRNHTTGYAATTSAKTAARHSFSLGFAQMFSKTRERQSPST TSSHNYFAGLRFDSLLFRDFISTGLSLGYSYGDHHMLCHYTEILKGSKAFFNNHTLVAS LDCTFLPARITRTLELQPFISAIALRCSQASFQETGDHIRKFHPKHPLTDLSSPIGFRSE WKTSHHIPMLWTTEISYVPTLYRKNPEMFTTLLISNOTWTTQATPVSYNSVAAKIKNTSQ LFSRVTLSLDYSAQVSSSTVGQYLKAESHCTF

542561 545401

CPn_0471 542561 545401
pmp_18-Polymorphic Outer Membrane Protein
TVONRSLSKSSFFVGALIGARTILIGATTILIANTPLSDYPDNQANQLTTLFPLIDTLTNMTPYS
HRATLFGVRDDTNQDIVLDHCNSIESWFENFSODGGALSCKSLAITNTKNQILFLNSFAI
KRAGAMYVNGNFDLSENHGSIIFSONLSFPNASNFADTCTCGAVLCSKNVTISKNQGTAY
FINNKAKSSGGAIQAAIINIKDNTGPCLFFNNAAGGTAGGALFANACRIENNSQPIYFLN
NQSGLGGAIRVHCECILIKNTGGVIFFNNFAMEADISANHSSGGAIYCISCSIKDNFGIA
AFDNRTAABROGGAICTOSLTIQDSGPVYFTNNQTWGGAIMLRQDGACTLFADQGDIIFY
NNHFHKDTSSHMUSVLCTRAVSLTVGACOGHSATFYDPILQRYTIQNSIQKFNPNPFHLG
TILFSSTYIPDTYTSRDDFISHFRNHIGLYKSTLALEDRAEWKVYKFDQFGCTLRLGSRA
VPOTTDDECSSSWUSVVININNLAINLFGILGNRVARKLWIRFTCGSSAPYGEDNNFIINL
SCHLUSLIDERNLDFYDTADLAGPILEVLLLUTVAKHINTDNEY POLLMTTOHYGYQG
WOPPWIFTITTSPT:SEDTVNTLHRQLYGDWTPTGYPVNPENKGDIALSAFWOSFHNLF
ATLBYQTDQGQIAFTAYSEATRLEYHONSNNDAKGFHMEATGYSLGTTDNTADNISFGVN
FOGLGSNLYELBERDDVASHTPTTAAGINNBAKGFHMEATGYSLGTTDNTADNISFGVN
FOGLGSNLYELBERDDVASHTPTTAAGINNBAKGFHMEATGYSLGTTDNTADNISFGVN
FOGLGSNLYELBERDDVASHTPTTYAHFILAYOPVLYQAMPEINVSLEDSSSWHHKARSYYGK
KPLYNLTYPLAAAGIGGSLAAMESPELHTTPFTOAIAVRSNUTAPOESGDVARKFSVH
KPLYNLTYPLAAAGIGGSLAAMESPELHTTPFTOAIAVRSNUTAPOESGDVARKFSVH
KPLYNLTYPLAAAGIGGSLAAMESPELHTTPFTOAIAVRSNUTAPOESGDVARKFSVH
KPLYNLTYPLAAAGIGGSLAAMESPELHTPFFTOAIAVRSNUTAPOESGDVARKFSVH
KPLYNLTYPLAAAGIGGSLAAMESPELHTPFFTOAIAVRSNUTAPOESGDVARKFSVH
KPLYNLTYPLAAAGIGGSLAAMESPELHTPFFTOAIAVRSNUTAPOESGDVARKFSVH
KPLYNLTYPLAAAGIGGSLAAMESPELHTPFFTOAIAVRSNUTAPOESGDVARKFSVH
KPLYNLTYPLAAAGIGGSLAAMESPELHTPFFTOAIAVRSNUTAPOESGDVARKFSVH
KPLYNLTYPLAAAGIGGSLAAMESPELHTPFFTOAIAVRSNUTAPOESGDVARKFSVH
KPLYNLTYPLAAAGIGGSLAAMESCHILTPYWHILLAYOPVLYQAMPEINVSLEDSSSWHLKSSTT
LAPNAAAAA

51-014 5455RI CFn_0472

NO TOBUST HOMOLOG PERSENT IN GENEDANK/EMBL AS OF LIT/1/98 EVFMAGTIKUSGLGKIPPKDMGDRSRGPSPKGELGSHEISLPPQEHGEGASGSSHIHS SSSFLPEDQESSSSSAASSPOFFSRVRSCVDRALKSFGNFFSAESTSQARETRQAFVRL SKTITADERRDVDSSSAAATEARVAEDAGVSGENPSQGVPETSSGPEPQRLFSLPSVKKQ SGLGRLVQTVRDRIVLPSGAPPTDSEPLSLYELNLRLSSLRQELSDIQSNDQLTPEEKAE ATVTIQQLIQITEFQCGYMEATQSSVSLAEARFKGVETSDEINSLCSELTDPELQELMSD GDSLQNLLDETADDLEAALSHTRLSFSLDDNPTPIDNNPTLISQEEPIVEEIGGAADPQR TREMMSTRUMNQTREALVSLLGMILISILOSILHRLRIARHAAAEAVGCCTCRGEECTSS HEEGMEN STEELVSLLGMILISILOSILHRLRIARHAAAEAVGCCTCRGEECTSS HEEGMEN STEELVSLLGMILISILOSILHRLRIARHAAAEAVGCCTCRGEECTSS HEEGMEN STEELVSLGMINTET TOMOLOGUM POVAFA FALL TAMA HERGAFTET TOMOLOGUM POVAFA FALL TAMA HERGAFT
CPn_0473 549602 548070
NO robust homolog present in Genebank/EMBL as of 11/7/98
GSIMAVGGVGGSRSPSPIPPNRRNSEDGKVSPKDNLGEHTVSSSDSSLASGGPTIEERKA
QLGGTDKIFLPSVKEPGDSQTSGRSGVLQRIWKGVKGVFKKTPQARPEVSSPRLPSHVQH
GQRLPGLEGFRDRIQKRSENPEADLGKMKRSYSDGDLDRVGHDSNEDSTEDSRSEGGEPS
SKSSSFLSGVRGAVSTVOGALGDIKGKFQRSASEDDLTTQGEDSAGDTVKERRSEEAEAS
SKSSSFLSGVRGATSTVOGALGDAKEKVSAFGEQAAGAIRSAPGNIRTRFQRSSEGDLS
NVNKAAKHLRKALENLEKVAPEQVSPEVASRVQSLLARMEQLTHQEPPTVEDLITFVESN
VGSDSVEYASIVPQDGSQAPAETAEAPETGGVEGSAAQGAWKALRDFVVSIFQAVASFFR
AIASRLSSARRESAVDDLASESNTQWFVEQEGVSNPSAAPSLSFAEEIARRAAEMSNRNA
QSLEKLESGNVTDPVIQQGLGLARSFAPEGQ

CPn_0474 551600 549807
CT365 hypothetical protein
LKIIISISFMSTSPISNDPRYLSLSNATEKTSLLANSRSLSPVPNSLVPSNPEDTGLRKS
IFTHSVTLFAGLVVLLVAVSVVVVALTVLAPGVPQAILLGIAISGVGIGGFSIMKSLVYM
VRDYMSPRMQESSRIKSALAVGTGFTVMGLVMKVGANFVPGGYGGLVGSLGSSAYSRGSQ
TTLASFSHYIYTKFFRSEKVAKGEKLTEAETIKEAKKLHYITLSIATIGVGLAVLGILLA
LAGTVLLGGAPATIAIILAPPLISIGLTTVLQTILHSSIGKWRAFLLTQEKKDLFVDTSL
KDIRLEKLPPSEVEESETSQSVIEVPDSEGIAETRISAEEIDTRLSLTTRQKVIFALATL
LLLASIAAFIVTGFGGLTVMQVLLVASVGSAVASVTLFMVSSGFSYVAYQLKARINISKL
KWKEAKNKKRVRQFLIESGVIASDREFNQMWKTVYKKQIQKTDAAIREEVRNFEKGGEVN
SALVGGILLGVGTGIMLLALVPAFAPIVPGILALGGSTLGIAGSILMRKFVNWLYDELVK
LYERRRNRRELLYGPESKMRSIATDLVVEALAASHDHLFDLDGPVDFIDVDVDIDGAA

CPn_0475 553850 551685
glsB-Glucan Branching Enzyme
PSMVDKLIHFWDLDLLVSGROKDPHKLLGILASEDSSDHIVIFRRGAHTVAIELLGELHH
AVAYRSGLFFLSVPKGIGHGDVRVVHQNGLLAHDPYAFPPLMGEIDSFLFHRGTHYRIYE
RIGAIPMEVQGISGVLFVLWAPHAQRVSVVGDFNFWHGLVNPLRKISDQGIWELFVPGLG
EGJEYKWEIVTQSGNVIVKTDPYGKSFDPPPQGTARVADSESYSWSDHRWMERRSKQSEG
PVTIYEVHLGSWOWQGGRPLSYSEMAHRLASYCKEMHYTHVELLPITEHPLMESWYYQVT
GYAAPPSRYGTLQFGYFVDVLHKENIGIILDWVPGHFPVDAFALASFDGEPLYEYTGHS
QALHPHWNTFTFDYSRHEVTNFLLGSALFWLDRWHIDGLRVDAVASMLYRDYGREDGEWT
PNEYGGKENLESIEFLKHLINSVIHKEFSGVLTFAEESTAFFGVTXDVVQGGLGFDYKWNL
WYMENFHYFMKDPMYRKYHKQHLIFTSLWYAFQESFILPJSHDEVNKGSGLVNLKJEDT
WTRFAQMRVLLSYQICLPGKKLLFMGGEFGQYGEWSPDRPLDWELLNHYHKTLRNCVSA
LNABYIHQPYLMMQESSQECFHWVDFHDIENNVIAYYRFAGSNRSSALLCVHHFSASTFF
SYVERCEGVKHCELLLNTDDESFGGSGKGNRAPVVCQDQGVAWGLDIELPPLATVIYLVT
FFI

CPB_0476 554877 553858
CT865 hypothetical protein
GGERRADWGDCMIDIMOHFKPYTMVFGQKLPIPGSLLYAQVFPTLWRLFSSKHEILNEQT
LQVGGPLKRFAVFQDLHRGGLAVTSERYKYYLLPSGECTQSIKGKLPSAAQAGPLLSLGV
KKHÄDWQKVRCRDLKEILPLWFRFAAWAPKGSYRDLETTAIGSLVKTAHQRVLHRETTE
LAPALLSIALAGFSECFLPRSYDEBFFQGILPDGDPEGGVPFELLSYSFGMIQDIFLRHQ
GQEVEILPALPPEFPCGRLHVALPNLGTLSIVWTKKTIRQVELHAEYSGEVFLKFCSSL
CSARLREWSERRLSGSKRLSLGETLEIKAGTTYLWDCFHK

CPD_EG477 556112 554844

*ydev_Bs Hypothetical Protein

RYM_FVAEVKGTFKLVCLGCRVNQYEVQAYRDQLTILGYQEVLDSEIPADLCIINTCAVTA
SAESSGRHAVRQLCRQNPTAHIVVTGCLGESDKEFFASLDRQCTLVSNKEKSRLIEKIFS

YDTTFPEFKIHSFEGKSRAFIKVQDGCNSFCSYCIIPYLRGRSYSRPAEKILAEIAGVVR
GGYREVVIAGINVGDYCDGERSLASLIEQVDRIPGIERIRISSIDPDDITEDHRAITSS
RHTCPSSHLVLQSGSNSILKRMNRKYSRGDFLDCVEKFRASDPRVAFTTDVIVGFPGESD
QDFEDTLRIIEDVGFIKVHSFPFSARRTKAYTFDNQIPNQVIYERKKYLAEVAKRVGQK
EMMKRLGETTEVLVEKVTGQVATGHSPYFEKVSFPVVGTVAINTLVSVRLDRVEEEGLIG
FIV

CPn_0478 557640 556210
hflx-GTP Binding Protein
whcGPLDTIDTPGEQGSQSFGNSLGARFDLPRKEQDPSQALAVASYQNKTDSQVVEEHLD
ELISLADSCGISVLETRSWILKTPSASTYINVCKLEEIEEILKEFPSIGTLIIDEEITPS
CQRNLEKRIGLVVLDRTELILEIFSSRALTAEANIQVQLAQARYLLPRLKRLWGHLSRQK
SGGSGGFVKGEGEKQIELDRRWYRERIHKLSAQLKAVIKQRAERRKVKSRGIPTFALI
GYTNSGKSTLLNLLTAADTYVEDKLFATLDPKTRKCVLPGCRHVTLLTDTVGFIRKLPHTL
VAAFKSTLEAAFHEDVLLHVVDASHPLALEHVQTTYDLFQLKIEKPRIITVLNKVDRLP
CQSIPMKLRLLSPLPVLISAKTGEGIQNLLSLMTEIIQEKSLHVTLNFPYTEYGKFTELC
DAGVVASSRYQEDFLVVEAYLPKELQKKFRPFISYVFPEDCGDDEGRGPVLESSFGD

CPn_0479 558434 55761b
phnP-Metal Dependent Hydrolase
AIGMYND 19585IGKLVFLOTGNPEGIPVPFCSCRVCONTGIHRLRSSVLIGYONKTLVI
DAGPDFRTOMLWAGVSELDGVFLTHPHYDHIGGIDDLRAWYIVTGRSLPLVLGASTYRFL
NKAKEYLFATPRVESSLPAVLEFTILNEDGOGEFFCGIPYTYVSYYQKSCHVTGFRRGNL
AYLITDLGGYDAKIFGVLDNVETLILLSAGPUSTFIFFCHRSSHLTVEGAKAFANHAGIKN
LITHIGHGLEAERDQHPEVTFAYDXMEVLWTL

CFn_0480 559375 558650
ST 38 1 hypothet Loll profein
DERIGWITGREGEPVOKKEQDCMLGSLPCYPGACHTEEYKNRYFYCQLCAEVVGPYVV
FVTVVDVOGAPPMILLOVLRCKOHRPCHEPVHGPTTSLWALEPVCKGAPQLESAMYELCS
QVRNPDTCSTVSWVFGGETFAGLTVGVMVEAPLTAGLSAWTECTTCXVGATLCLFATL
MAYLGR HVREWLNIGHEYTTQCHCRQTQAHSQNYCVTTEYPATCALSQPTTKLPNGSRR

•

CPn_0481 5.6.973 559339
No robust homolog present in Genebank/EMBL as of 11/7.98
SCLRIEGILMATSYPYTSSTS/GEANSSNERFTERTSRMYYAAL/UGALSCLIFIAMIVI
PPOVOLWAVVLGPALGCLLISLAIVFAVSGLVLGKTLEPSREATPPEIVAQKEMTTOQDV
LGNEYWRSELISLFLRGDLHESLIVDSKDRSLDIDQSLQNILKLEPLSTTLSLLKKDCVH
INII LHLVRQWNLLGYDLGFEVTAHAEELLLFLIEEQYYSPDILKLIRYGDALQATSPLM
WWW.1520.WF.A. 1520. 1

CPn_0482 561764 560961
artJ-Arginine Periplasmic Binding Protein
NLAYRAQTFMIKQIGRFFRAFIFIMPLSLTSCESKIDRNRIWIVGTNATYPPFEYVDAQG
EVVGFDIDLAKAISEKLGKQLEVREFAFDALILNLKKHRIDAILAGMSITPSRQKEIALL
PYYGDEVQELMVVSKRSLETPVLPLTCYSSVAVQTGTFQEHYLLSQPGICVRSFDSTLEV
IMEVRYGKSPVAVLEPSVGRVVLKDFPNLVATRLELPPECWVLGCGLGVAKDRPEEIQTI
QQAITDLKSEGVIQSLTKKWQLSEVAYE

CPn_0483 561830 564964

CPh_0483
No robust homolog present in Genebank/EMBL as of 11/7/98
IILIKKRAIFERMFPIPPPHCPPNNKNNFYHLTTDTKDPLLLRILRTIGYVLLHIITLGL
LLLHYYKHHRVVRKEGLPTPPTLDKGEPERTIELAKOPFKDGEDKKPDVPKROTPPPED
TPPPPKAPSPASPKVPKQPADKKRTPPPEAPPPPVRVATFMPLRPSSQCYWQCLNRWS
MVLRRAPLPLPAMQVDPILGDFNPHFVASYPNRIDNEPMYFQIKQFKKIAQNPDLPQOHR
RLAQLSLEQALYLNDNYYLUNVEGIGNCFYRAYAVGWLSALYEESSRNDIVFEQEATRLL
DLPFASSSPANANLCAEMAELLQLCSTYCSFIDLYDGVILSQKHTATLIAFLKRLSAYAI
RQQIAASSNEETARALFISDMQDDLLPSVLEFLAANRPYSELFQNLIDHSALPYMQSRDK
LFLLLEHLPALFITDAELQKMSPEDQQLRKQYEREIREAFAKLSRRIADSGMTERFNAI
VKDHLPEAIRCQYSRFLATIENRSGDLPWSPALSFFAFLCTCPSVRFHKLCATFYKSLE
DIIIASAPPQRSIQEILQISNASLSYLNEDLDSSWQREVISSNIMTILTHESLTLESSM
PQLETLHKRIANLLKNVISTSFETPPLSNQPDLLSNLVNKLLVAIHSKLELKEHFNTVG
ARSLRLTRDEGSGLSQEQDLLYTQAVQLLFFILQHPQVNNRPETKDAVKELKMLLPFLQ
YAFKKVENEKKLQKLLRSILGSLVLKPPARYPSTPSNKDKETFCKFWSRHPEVMVLDPIL
EKNCMGFLRATFPNYQLETEAILLEKEIESTFRNGWNYFLTRLNLFGSKLGSPSSPTALS
DGFSKSFLIFCFLNNYPKLLQKKTPLAARLDAFQREASHRFTQVKDKLLLSLKYGFPLAT
ATINQYSRARDQLICNLLKNTVTASDGFCRSGFRQSLIGYLHSLSSNELGDILDDVKQA
EANDVAAMTTVPLQPFAVCLIMSDRDTVSEENIENFVAMHGFLNTISPERDARIFLIRFP
NHYGCLLPRNPRTEDQNSKPDSSNP

CPn_0484 564931 565824
arod-Deoxyheptonate Aldolase
RSELKTGQLKSLVLHEVLILTFTYPLPRTLKQHPDEVHTVPISPNLSFGEGSPILIAGPC
TLESYEHTVSSALTVKEAGAQVFRGSIRKPRTSPFSFQGWEKECVLWHKEAQSIHGLPTE
TEVLDVRDVEITAEHVDILRIGAKNMHNTPLLQEVSKSHRPIILKRSPAATLEEWLCAAE
YILASSPSCPGVILCERGIRTFEHSTRYTLDLNTVALLKEISSLPVIVDPSHAAGKRSLV
LPLASAGLSVGADGLMIEVHAHPEKALCDAKQQITPEELHLFAKKHFCPSESRAHAIS

CPn_0485 565993 566229
CT382.1 hypothetical protein
OPIGRTPTRVFLWRFMIKQACKFYLLQCLLCALYWLLKYCRKLLKGTLHHSEETLYQALL
SSLIDLLYQLKQLPAPTNE

CPn_0486 567799 566405
hypothetical proline permease
aQHRSLLKGNIFHLCGGVLYFMNFSLFLFFLIAIQGICLYVGRRGSKKVEDRESYFLAGR
SLKIFPLMMTFIATQIGGGVLLGAAEEAFCYGYGGILYPLGVALGLIFLGMGPGKRLAEG
SLTTVVSIFEVFYGSKKLRKIAFLLSAGSLFFILVAQVIALDRLFSSFPFGKYVTVAFWI
VLASYTSTGGFRGVVRTDVIQAGFLLIAVLVCGVSVWLSVPKSLSVLDPFQSLPCAKLSN
WIFMPMLFMLVEQDMVQRCVAASSPKRLQWAAVGAGLVLLLFNFIPLFLGSLGAKAGLKA
GCPLIDTIAYFCNPSLAAVMAAAIGVAILSTADSLMNAVSQLIAEEYPTLKAPYYRYLVL
GLAVAAPLVAIGFTNIVDVLILSYSLSVCCLSVPVCFYLLAPKGRRVSGAAAWAGVLVGA
LGYGWVQIVSLGMFGELLAWVGSLVAFSFVGFIEITWKNKVKTQT

CPn_0487 569833 568112
CT384 hypothetical protein
RTGGISLTYSSFRWASFRCYSLIFFCFCGSLFGSESLRYQLLIQDFAKVSEEGIGLLES
KEYSLLQAKLVLRALAQNSSFDDWFRSFKKCQISYPELAHDRDVLLEFGIQVLREGIENP
SVTVRAVSVLAIGLARDFRLVPLLLQSCNDDSAIVRSLALQVAVNYGSESLKKAIVELAR
NDDSIHVRITAYQVVALLQIEELLDFLAERAENKLVDSVERREAWKACLELSSQFLETGV
AKDDIDQALFTCEVLRMGLPETTEIFTELLSVEHPEVQESLLLSALAWSHQLQNHKEFL
SKVRHVMCTSPFAKVRFQAAALLHLHGDPLGRDSLVEGLRSPQFLVCEAASAALCSLGIH
GVPLAKEHLESLSSRKAAANLSILLLVSREDIERAGDVIARYLSNPEMCWAIEYFLWDAQ
WNLRGDTPFLYSDMIKREIGRKLITLLAVARYSQAKAVTATFLSGQAQGWSFFSGMFWE
EGDVKTSEDLVTDACFAAKLEGALASLCQKKDQASLQRVSQLYNDSRWQDKLAILESVAP
SENLDAVPFLLDCCHHEAPSLRSAAAGALFSIFK

CPn_0488 570147 569767 hita-Hit Family Hydiolase RKLPTCFAVNVTRSRDHMTVFKQIIDGLIDCEKVFENENFIAIKDRFPQAPVHLLIIPKK PIPRFQDIPGDEMILMAEAGKIVOELAAEFGIADGYRVVINNGAEGGQAVFHLHIHLLGG

CPR_0489 571037 57009CT387 hypothetical plotolin
RIVPALENYPOLVTNEDMERRIVPMOTERSIC/THOGSPHADEVTACALLIIFDLVDENKI
IR:RAPAVVLSKCFYXCOVGXYS LENRRDHHOVGYD SWISSA: MILHY LKREGYMOCEE
HFELMRUCHGODEOIM REPEX POJECOS DI I FLYNYP EERETNISDADFSC ALHFTIDE
LCKLPYKFOYDRVCRUTVRIAMETEDMETLYFDPPLAMQENFFILXSEKHDAAFVCFPSCD
OWLLIGG I FONLDRRHDYPV PEPEHWAGLIG KELGKVSGT PGAVFCHKGLFLGVWTNREGC
ORALDETEDDWETE

CPn_0490 5 1277 573333 CP387 hypothetic (Lipiotzin LMYNDLIIMIBDAADTKORI ASHLIPPLEMI LYYSEVLIENI PAYFIXIFHLIPSOCIOVNEKS SLADUGVAVUMHISILINKARRE APLIPTIEM, SUPTATAMLEDLEESSYVCKI MAADDRRI VRSBCYLMEMPTHETORYSI CLUPPSKYLISHI TILBELIIMPUVVELE HISTI CYTITETYS FLPLM:R:JLTRPHLKIRKFLPLYGMVTDRPPVPEDHKILLIKTEPLHIRTVFARVVQDLL
PQGLRHTAADILEPTTQESGDIYEFYGSTSEPIERIPLEFFTLEPYKEHSFFFYRDMLQE
TLEOPQEVFRVFESIPEGEDQAAMFISKGSELLELSQDSWIIKRRISPSDERHAREIQKH
IEDQPCFPPLKAMETOHITSGGVLFGRYFPSASIKGMFLSNYSRYYLQHIYFQIPSFTSG
EFFSNRDRSFLLDLYFAGISVFWADLESKRLLQYIKRRNKDVGMFVPKHQAEQFAQSYFI
GIHGSCLIAGDYDEFLRELLTGMHTLSQQFTIPEFPPQTPLAILTGGGSGAMELANRVAT
ELSTLSCGNLISLDTINAYVEXKMSYAIPDLLERQADFHVDLAVFVIGGMGTDFELLEL
ISLKTGKKALVPVFLIGPVDYWKSKITALYNSNHAVGTIRGSEWVHNCLFCLSSAKAGIA WINDLE GERRYPERGEV

574595 5/3336

CPn_U491

774595

773336

CT389 hypothetical protein

ILSSLYTVFTMKTAFHSCYSWFCMLFSFLVLFVGGIAGGEPLCPDCKYETKSVLRSDQLP

DHLWNYENDCYLTGYVQSLLDMHFLDSRTQVVIEKNRAYLFSLPVDSSLSEAITNFVRDL

PFICAVEICERPYGECITRSSAERPLLPKEKTLGMPIFCGKEGVWLPQNTILFSPLIADP

RQVTNSAGIRFNEKVVGNRVGATIFGGDFILLRLFDVSRFHVDCDFGIGGGVFSVFDLDH

PESCHVNSDFFVAGLWSGAIDKWSFRFRLWHLSSHLGDEFILTHPNFPRFNLSDEGVDLF ISFRYTPQIRLYGGCGYIVSRDLTFPERPFYCEWGAELRPFGLREGNLHAQPIFAMHFRC WEEQKFGLDQSYILGMEWAKFQEIGRKIRAVLEYHQGFSKEGQFIREPCNYYGFRLTYGF

CPn_0492 574643 574804 No robust homolog present in Genebank/EMBL as of 11/7/98 LFSLIFPICEERNSQQTYKHLHVESACFLLESPLKIHWSSPYGFPPFYRRDLKL

CPn_0493 575142 574855 No robust homolog present in Genebank/EMBL as of 11/7/98 SKTEGSHSKTSKGFVGRFVQWIRTFTGRGSKKRSPSSFSPTHPYIRLRTYTRSPKQSGVE RKQEDAETSFIETPKGILKKPGNKDPKGKHVHWKDS

575370 575146

VIMIRVNPYGSYRGRNPSPEDGKKDVPLSGNSRLHRRGGIRRKHKSASVGVTSGSKTGKA SLEKKVKGISEAHFK

575507 576793

aspC-Aspartate Aminotransferase
RRLKKNQKMAIQKAGAFLRCLPSESRPYLEHAMRRNPHFSLLKPQYLFSEISKKLAQFRK RRLKKNOKMAIQKAGAFLRCLPSESRPYLEHAMRRNPHFSLLKPQYLFSEISKKLAOFRK
ENPEISVIDLSIGDTTQPLCRSITQAIKEFCVSQEKQETYRGYGPETGLEKLRTKIASEV
YENRISPEEIFISDGAKPDIFRLFSFFGSEKTLGLQDPVYPAYRDIAHTTGIRDIIPLAC
RKETGFIPELPNQQSLDILCLCYPNNPTGTVLTFQLQALVNYANQHGTVLIFDAAYSAF
VSDESLPKSIFEIPEAKYCAIEINSFSKSLGFTGMRLAWNVIPKELTYDNNEPMINDMKR
LFÄTTFRGASLLMQEAGYYGLDLFPIPPAISLYLTNAQKLKKSLETAGFSVHGGDHAPYL
WWELPEGISDEEAFDFFLHQYHIAVTPGHGFGSCGQGFVRFSALTQPQNIALACDRLCTA
SLKETMVLA

CPI_0496 576751 577812
CPI_0496 576751 577812
CTS-07 hypothetical protein
PPEYRFTKRNDGSCMTILRKLSQYLFFFSLFCSFIYVATCGSQPDSVSSPKIAIFLSFPH
PLIEDCSKSCIETLKDFENLPEIVVLNAEDSIVKARKIARSLHTDKNVVAIVTLGTIATK
WEHLETCKPKYIVAAVPDRESLTLPKNTKNIYGNDFLLDINQYCFAIGAVATNAGSIVYL
KPSEPFPSDLQKEIVKKLHASGIEVIEISITSSTFKTRIRQAIDKRPSAIFIPLSPLSHK
EGTAFLQEILKEKIPIITDDTSLISEGACIACSVDYKKSGKQIAKIVHHLLYNNHDVDSL
RKIIAQRISPTTFNEDIIKYLGIKLHKTERNQFLSFKSKKLEKSEKGKNVAVS

10 10 27 578107 577820
CTJ388 hypothetical protein
IFORVVLDDSWILEVKVTPKAKENKIVGFDGQALKVRVTEPPEKGKANDAVISLLAKALS
LPKRDVTLIAGETSRKKKFLLPNRVQDIFFSLHIDV

579062 578085

CPA_U498 579062 578085

No EDDUST homolog present in Genebank/EMBL as of 11/7/98
YCRERAPFMNRKARWVVALFAMTALISVGCCPWSQAKSRCSIDKYIPVVNRLLEVCGL
PEÁENVEDLIESSSAWVLTPEERFSGELVSICQVKDEHAFYNDLSLLHMTQAVPSYSATY
DCAVVFGGPLPALRQRLDFLVREWQRGVFFKKIVFLCGERGRYQSIEEQEHFFDSRYNPF
PTEÉNWESONRVTPSSEEEIAKFVMMQMLLPRAWRDSTSGVRVTFLLAKPEENRVVANRK
DTLLLFRSYQEAFPGRVLFVSSQPFIGLDACRVGQFFKGESYDLAGPGFAQGVLKYHWAP
RIGLETLAEWLKETNGCLNISEGCFG

CPn_0499 580404 579205
No robust homolog present in Genebank/EMBL as of 11/7/98
LSVYLLIFYFCNCSTMSSVNQSSGTPNPEEVTSPESTEENKNVSSDEAQAFHAVALPIV
TOLSLPEGVGTSSEETASNPRVDEDIVAEVSSSRAVADQISSLVERVGELLDDLKGAQSLF
TSFQSELKNCLPAWKSSTRRLETRGAGDNADIARLELFRSDYEAVLGHANQFHGKAHLIL
SKLTDVHHKLQGLGREDLSLAFDNNDRVLEHLGSLGLDVDAEGNWSLSCERGIPRLVLTA
SMLVQIKKVNLPTVEELRTLQGTTESSSDPRVESLSCCERLLNELRLWANFVGFISS
CYDNIVFVLMWIVRRINLLPGLGCLPFHNPDASQEDQRSSSGERSTRRERLSRRSDLSEE
EMIVRAEGESIHPESPHGDGRNQPSRGDKQDSDSEEETEL

580647 582362

CPn_0500 580647 582362
pros-Proly1 tRNA Synthetase
QPHSMKTSQLFYYTSKNANKSAAVLSNELLEKAGYLFKVSKGVYTYTPLLWRVVSKMMNI
IREELNAIGGQELLLPLLHNAELWQHTGRWEAFTSEGLLYTLKDREGKSHCLAPTHEEVI
CSFVAQWLSSKRQLPLHLYQIATKFRDEIRPRGLIRSRELLMEDSYTFSDSFEONNEQY
EKLRSAYSKIFDRLGLAYVIVTADGGKIGKKGSEFGVLCSLGEDTICVSGSYGANIEAA
VSIPPQHAYDREFLPVEEVATPGITTIEALANFFSIPLHKILKTLVVKLSYSNEEKFIAI
GMRGDRQVMLVKVASKLNADDIALASDEEIERVLGTEKGFIGPLNCPIDFFADETTSPMT
NFVCAGNAKDKHYVNYMDRDLLPPQYGGFLLAEGGDTCPENDGHPYRIYQGIEVAHIFN
LJTP/TDSFEVNFQDEHOQTQQCWMGTYGIGVGRTLAACVEQLADDRGIVWPKALAFFSI
TIAFNGGDTVSGELAETIYHELQSGGYEFLLDDRDERLGFKLKDSDLIGIPYYLILGKSY
QSGGIFEIESRGGEKYTVSPEAFPTWCQNHLA

CPn_0501 582424 583650

CPG_0501

Figh-HTH Transcriptional Rediessor

ILLTEGOOPPIMESVTIVLVOLPMARSKVSKPDSKILDILFATTELYLKTOQPVOSKTLK

ESPECIDLETEAT IRNYTABLEARIGELKKRIET XORFIPTELARRHYVDHIQEESPERAETISAPI

FORKTOGLEESINKI IKHORKATELLGE ILLDPTEGOOPPEPRODSYTHO (ITVOVKQRAVT

ILTTEFFOTFTOTAMLERACTOLITIKRIEKTLONY IPPLPTNEELSKKEEHLOMOLYNEV

VVRYTTRYONFSSEDLYGT MYKLLKYLAFKSPEVLAGOSSEPBARROM, ELLNIGMIKG

RATTAFICKELSSI ILTTENDERSVITTI PYYMNRSPEJAAG ILGP INLTYFRALFILLKLAN

KINETUTOSEYKPKLSIPPREMENCKLONFFILRTEYGGTKLLPSKTTI.

584201 CPn 0502 583641

GDVMTDTPPENEEQHESN/ONENEVEHLQQEIVTLKTELKEKNDKYLMALAESENSRKRL OKEROELMOYALENTLIDELNPIESMEKALGFATOMSDDVKNWALGFNMILNOFKQIFEE KGIIEYSSIGOKFNPFLHEAVOTEETSEVPEGTILEEFAKGYKIGERPIRVAKVKVAKAP

CPn_0503 584225 586213

ige set de t 1 g t tal 1 Tollie PAKROAVTNPEKTLGOTKPFIGRKYGEVAGEIGTVP/TVTSGSKGDAVFEVDGKQYTPEE PAKRQAVTNYEKTLGITKFF IGNKYJEVAJELJTVP/TTVTSGSKGDAVFEVDGKQYTPEE
IGAQILMKMKETAEAYLGETVTEAVITVPAYFNDSQRASTKDAGRIAGLDVKRIIPEPTA
AALAYGIDKVGDKKIAVFDLGGGTFDISILEIGDGVFEVLSTNGGTLLGGDDFDEVIIKM
MIEEFKKQEGIDLSKDNMALQRLKDAAEKAKIELSGVSSTEINQPFITMDAQGPKHLALT
LTRAQFEKLAASLIERTKSPCIKALSDAKLSAKDIDDVLLVGGMSRMPAVQETVKELFGK
EPNKGVNPDEVVAIGAAIQGGVLGGEVKDVLLLDVIPLSLGIETLGGWTTLVERNTTIP
TQKKQIFSTAADNQPAVTIVVLQGERPMAKDNKEIGRFDLTDIPPAPRGHPQIEVSFDID
ANGIFHVSAKDVASGKEQKIRIEASSGLQEDEIQRMVRDAEINKEEDKKRREASDAKNEA
DSMIFRAEKAIKDYKEQIPETLVKEIEERIENVRNALKDDAPIEKIKEVTEDLSKHMQKI
GESMOSQSASAAASSAANAKGGPNINTEDLKKHSFSTKPPSNNGSSEDHIEEADVEIDN
DDK

CPn_0504 586418 vacB-ribonuclease family 588514

VacB-ribonuclease family
ATOFTSETTGFLVQCPKLTGGAQLLKKPKRKPGRRTYGKSLKIFIPGTLFVHARKGFGFV
SPDNPESYPPDIFVPARDLRGALDGDHVIVSVLPYPRDGQKLKGTISEVLARGKTTLVGT
ITSLVSPTSALAYTSMSGSQSLIPVELLPGRTYKIGDRILLSTPPWVDKPQEGASPALQM
LEFIGHITNAKADFQAIQAEVNLAEEFPPEVIEEASLFSQKHITQVLHSRKDLRDLLGFT
IDSSTARDFDDAISLTYDHNNNYILGVHIADVSHYVTPHSHLDKEAAKRCNSTYFPGKVI
PMLPSALSDNLCSLKPNVDRLAVSVFMTFTKSGHLSDYQIFRSVIRSKYRMTYDEVDNII
EKKHSHPLSKILNEMATLSKKFSDIREERGCIRFVLPSVTMSLDNLQEPVALIENKGTFS
KKLIEEFMLKANEVVAYHISHQGVSLPFRSHEPPNDENLLAFQELAKNMGFDITFTPTQE
PDYQYLLQTTSAGHPLEQVLHSQFVRSMKTASYSTENKGHYGLKLDYYTHFTSPIRRYID
LIVHRLLFNPLSIDQTHLEIIVRACSTKERVSAKAENSFENLKKTRFINKFLQEQPKTTY
HAYIITANHEGLSFVVTEFCHEGFIAAAELPKEYSLKKNALPESIPDKMKRGASIKVTID
SVNLLTOKIVMSTATTEDKYPKK IKKTPSKKKGTKKRAS SVNLLTQKIVWSIATTTEDKPKKIKKTPSKKKGTKKRAS

CPn_0505 588471 589106
*3-methyladenine DNA glycosylase
RKRLLRKKERKKEPRNVLOEHFFLSEDVITLAQQLLGHKLITTHEGLITSGYIVETEAYR
GPDDKACHAYNYRKTORNRAMYLKGGSAYLYRCYGMHHLLNVVTGPEDIPHAVLIRAILP DOGKELMIORROWRDKPPHLLTNGPGKVCOALGISLENNRORLNTPALYISKEKISGTLT ATARIGIDYAQEYRDVPWRFLLSPEDSGKVLS

CPn_0506 589055 589840
CT421 hypothetical protein
CPMEISPIPRRFGKSFILNNIKLYSKETNAHFLISCRRIMKKYFITGLVILLPLAITIAI
VTMIMNFLTQPFVGLASEFFEKFSFYTKHRALLKFVLQIILLFGLFFATVLLGFLTRIMI
FKSLLSIYDKILHRIPIIKTVYKAAQOVMTTIFGSKSGSFKQVVMVPFFNANVQCIGLVA
GDAPTVCCTGEREDDPLVTVFIPTTPNPTSGFLTLFRKSDIVFLDMKIEDAFKYIISCGV LSTPMACPSSPLPDELHQDQGS

CPn 0507 589898 590122

CT421.1 hypothetical protein STPYPQFPLSGEIKKFNIELFMTRMSKQARRRAKSPKKRKPKYAIVHPAPAPRIVYKLHT

NALSTSDSIFIPKIG

CPn_0508 590133 590300 CT421.2 hypothetical protein SRIMSRHRSYGKSVKGVTKRNVLKRFERVEVLRKLGRWNDSTAKKVTGLPKTPILK

590299 590808 CPn_0509

(Predicted Metalloenzyme)
NKFYFLYGNFIRVTQEKIKIHVSNEQTCIPIHLVSVEKLVLTLLEHLKVTTNEIFIYFLE
DKALAELHDKVPADPSLTDTITLFIDAFGDPAYPHVLGEAFISPQAALFFLENTSPNQED
IYEEISRYLVHSILHMLGYDDTSSEEKRKMRVKENQILCMLRKKHALLTA

590804 591973

CPn_0510 590804 591973
tlyC-CBS Domains (Hemolysin homolog)
QLNMLHILLAIFCILLFLAFGLTQPSCHGSSKFLKTLNQRFFKDKGREYPPFPSAPTILA
TLLCILYGALGTKLYTLLPPKTAHKDLLFWPLYSLSALIAYGFLPPWISTKVPKETTAHL
RFLASVFQLGFPPLGLFYRRRPMQVWRSSTSFQSCLSEALISAFFOLIVREWIMPKVDIF
ALPEETTLQEALVLVSEEGYSRVPVYKKNLDNITGILLVKDLLLLYTSSHDLSQPISSVA
KPPFYAPEIKKASSLLQEFROKHRHLAIIVNEYGFTEGIATMEDIIEEIIGEIADEHDVQ
ENTPYKKIGSSWIVDGRMMISDAEEYFNLKIDHENSYDTLGGHVFHKVGAVPQKGMRIHH
ENFDIEIITCTERNVGKLKITPRKRKFNIS

592141

rsbV-Sigma Regulatory Factor
MSDIQKEEHGSTTIFHLHGKLDGISSPEVQENICQSLAAGSKNIILDCAHLDYMSSAGIR
VLLQSYHQVGQHSGKIVLTTVPKTIEQTLYVTGFLSYFKIFNTVDEAIQTLNKDGD

CPn_0513 59 For 3 exidereduct ase

CDE SCREUGEAVERGIBERMAN, GODDING TOUTRIBITEAH IARGCEBBUNDAARQUE TERKTDVOEDITRORENO TYYKKERDAERO JAATOHITEE IAHOGAHAMALCTBOBEKARABOTA AND CANABITRORENO TYYKKERDAERO JAATOHITEE IAHOGAHAMALCTBOBEKABOTA BATOTO OMER 1987

TKIKEYDEO HIKALTAIEYAYLSDLDNLSIRDVLLTLKDAGLDSIPGGGAEILVDKIRN FLAPKRESSOPENTHKMARQLGTHSNTTMLEYHKEGPEDLVTHMYKVRDLQDETQGFKN FILLKFAQENNVLGKRERKSGQGHAIPLKSLMAVARIFLDNFSNMKALWNYLGIEAALDL LSCGANDLSSTHMGEKVFQMASSKEPIKMDAEGMAALITQQGRTPCLTNSSHV

CPn_0514 595690 596520
CT427 hypothetical protein
GNGGPHHTTRENAMGHOLOPCISLOCVGYINSFPLSLOLIKRNDIPCVLAPPADLLNLLI
SHGLEYALGEBER SERVER SER ASGWYDLTKLPFVFALLLHSTSWKEHPLPNLAMEEALQQFESSPEEVLKEAHQHTGLPPS LLQEYYALCQYRLGEEHYESFEKFREYYGTLYQQARL

CPn_0515 596450 597181
ubiE-Ubiquinone Methyltransferase
EKNTTKALKNSGNIMEESTNKPDCKKIFDSIASKYDRTNTILSLGMHHFWNRSLIQILGS
GYSLLDLCAGTGKVARRYIAAHPQASVTLVDFSSAMLDIAKQHLPQGSCSFIHSDINQLP
LENHSYPLAAMAYGLRNLSDPHKALQEISRVLMPSGKLGILELTPPKKTHPTYSAHKLYL
RAVVPWIGKSVSKDPDAYSYLSKSIQQLPKDHDLEDLFSKSGFYIAKKKKLFLGAATIWL

CPn_0516 598904 597255

No robust homolog present in Genebank/EMBL as of 11/7/98

RISISFRVSWFVKIILAVLGRAIAKAYYVCMVARGLCDFPTLVPNERLPIGPFFVPOHTS
GAKGKEFAKRNFSIISGLDILKLCILQRRPFALQWDNLSVKSDYEEAGPAIGIRSLEPQ
VSQISPAHGRLCSTLVQWAPILGSEEQLVWLEETMKRLKFPKSLGSKDAVIVDSEMYPVN
ANPTQEIPAASETVESSPVAPGNTTDTMPAASGTTDTTSGVSEAAAAEAAVDSTPGTEEE
PSFSLRYALVVQNVPYPEPPKEPEVMFTDEKKSLILEATRARRMELDLYNGYLADYELSK
DEIQKHVPDLPENWRTNWRMSERLYKFFFKKKEGLEEIFLNKEIGNMILARGLAATQSQ
ARIKVFNSLVAWLLQSFNVGRSCTAKPLPTSKLDLFKSEFESKPKNNILTEFLVASDEEI
LFKGLRVLEPGIEGWYDHPDQAGEIRSVLEGLVQAGRISGYWENQPFGRFVLRGVERRT ELVELLESLVASGEIMQFFESSDEEGAF1IDNEPSKTAMLKQRFKSCVRTKLVGSFADES LPRGRFTILV

599637 598795

CMT_U517

No robust homolog present in Genebank/EMBL as of 11/7/98
FIMSSLLSCGRIEPTRVTCSLKTYLEDTSQNQLSTRLVRASVIFLCALLIILVCVALSSL
IPSIMALATSFTVMGLILFVMSLLGDVAIISYLTYSTVTSYRQNRAFEIHKPARSVYYE
GVRHWDLGRSSLGTGEIPIVRTLFSPFONHGLHALAAKIFLFMEHFSPEPPMEPLUWA
CLIRDFRPHVSSLCFVIEKQGSSLRTKEGNTICEAFRSDYDAHFAMVDCYRLIHSKLIIE KWSLKNIDI I PSVMVREDYPSRPGEGYREGLLRMYGGKGAL

CPR_0518 600806 599832
CTM29 hypothetical protein
FMGTTYPVPQNPLLLRILRLMDAFSKSDDERDFYLDRVEGFILYIDLDKDQEDLNKIYQEL
ENAERYCLIPKLTFTEVKIMETFINEKIYDIDTKEKFLEILQSKNAREQFLEFIYDHE
ABLEKWQCFYVERSRIRIIEMLRNNKFHFVFEEDLDFTKNVLEQLKIHLFDAKVGKBITQ
ARQLLSNKAKIYYSNEALNPRPKRGRPPKQSAKVETETTISSDIYTKVPQAARRFLFLPE
ITSPSSITTSEKFDTEEEFLANLRGSTRVEDQLNLTNLSERFASLKELSAKLGYDSLTG DFFGDDDEKVVTKTKGSKRGRKKSS

CPR_0519 601707 600904
dapf-Diaminopimelate Epimerase
QPFKLRILVYMMAFYSPSTISKYFIYSCAGNRFLLGETLPEVEDVRFLCQETRVDGFLYL
KPŠSCADAQLIIFNSDGSRPTMCGNGLRCAIAHLASQKGKSDISVSTDSGLYSGYFYSWD
RYLVDMTLADWRASVHRLESRPDPLPKEVVCIHTGVPHAVVILPEISTLDLSILGPFLRY HÖTFSPDGVNVNFVQILGHCQLRVRTYERGVEGETAACGTGALASALVVSNSYGWKESIQ IHTWGGELMTVSQNRGRVYLQGSVTRDL

602233

CPHE-0520 60223 601646
CIPP-CLP Protease
ERMYFMADGEVHKLRDIIEKELLEARRVFFSEPVTEKSASDAIKKLWYLELKDPGKPIVF
VINSPGSVDAGFAVWDQIKMLTSPVTTVVTGLAASMGSVLSLCAAPGRRFATPHSRIMI
H@FSIGGPITGQATDLDIHAREILKTKARIIDVYVEATNQPRDIIEKAIDRDMMMTANEA KDEGLLDGILFSFNDL

CPHS 0521 603803 602241

glipa-Serine Hydroxymethyltransferase
kSLLKVFEKFKKFAIVEIFTKVVAVVSLLHKFLENASGKKGQSLASTAYLAALDHLLNAF
PSIGERI IDELKSORSHLKMTASENYSSLSVOLAMGNLLTDKYCEGSPFKRFYSCCENVD
AIEWECVETAKELFAADCACVQPHSGADANLLAVMAILTHKVQGPAVSKLGYKTVNELTE
EEYTLLKAEMSSCVCLLGPSLNSGGHLTHGNVRLNVMSKLMRCFPYDVNPDTECFDVAEIS
RLAKEYKFRVLIAGYSSYSRRLNFAVLKQIAEDCGSVLWYMDAHFAGLVAGGYFVDEENP
IPYADIVTTTTHKTLRGPRGGLVLATREVESTLNKACPLMMGGPLPHVIAAKTVALKEAL
SVDFKKYAHQVVNNARRLAEFFLSHGLRLLTGGTDNHMVIDLGSLGISGKIAEDILSSV
GIAVNRNSLPSDAIGKWDTSGIRLGTPALTTLGMGIDEMEVADIIVKVLRNIRLSCHVE
GSSKKNKGELPEAIAQEARDRVRNLLRFFLYPEIDLEALV

CPn_0522 603825 604655
CT433 hypothetical protein
REPLSPEKTSLAFKVKNVNQRMIKKNQGKKKNYFQYIPLKVQKLROPSFYPKRLMTLYLG
LNQKTARKYQAHYLPILTLFPYAKSTPONKRALQFLPQATHVILTSPSSTHLFLSRMTSL
LSKATLKTKTYLCIGESTKERLLSFLGOVKYVVATQEIAEGIFPLLQALPSSARILYPHS
SLARPVIRFFLYNRFPFSYPHYTVKPRKLKKNILSXYKKIIFTSPSTVRAFAKIFPRFP
EKTYWCQGRMTLQEFCKFSSQKQVSLLETLGKSRTSP

CPn_0523 604720 605052
No robust homolog present in Genebank/EMBL as of 11/7/98
FMACSATEGEOTAPSLEPPATRPRYNFKLALEVTIAIALVWIALIATTIAIGLCIHPLC
DETECTATEGETERSTICCHYARNYYIALDVVPDHCHLODMRSHSFIEDDR

CPn_0524 605079 606179

ID TOBUST HORIOLOG PROSENT IN GENEBARZEMBL AS ST 1177798

FZERKEMMOGERETE ROVINLEZVERDKETARKZETTAK KZELILAGIALGALVAG
ISSTETIZARVELAL TETALIZAVETELVERDMEZVESOMMOZVELORPKELIKAMOGKN

JUZZSDEMOFYNNILLI TETALIZAVETELVERDMEZVESOMMOZVELORPKELIKAMOGKN

JUZZSDEMOFYNNILLINEKEKVA IOTDASOFFORFETZSLRVI I KNOSTGI I FNEVOPENL

IGITATHILITI I LYTELKOK VWODY KORBOGPAKOFFOFFORVRVAKLENBALDOTENL

IGITATHILITI I LYTELKOK VWODY KORBOGPAKOFFOFFORFICKAR I ESHAA I VALPLEF

TOZZEVERE I LITETI SHVCORKBELPNOGEYRGA LLAYFNICKAR I ESHAA I VALPLEF

TOZZEVERE I LITETI SHVCORKBELPNOGEYRGA LLAYFNICKAR I ESHAA I VALPLEF

TOZZEVERE I LITETI SHVCORKBELPNOGEYRGA LLAYFNICKAR I ESHAA I VALPLEF

TOZZEVERE

695510 507283

CT398 hypothetical protein CT39M NYPOLNETICAL PROTEIN
GIIFMHDALLSILAIQELDIKMIRLMRVKKEHQKELAKVQSLKSDIRRKVQEKELEMENL
KTOIRDGENRIOEISEOINKLENQOAAVKKMDEFNALTQEMTTANKERRSLEHQLSDLMD
KQACGEDLIVSLKESLASTENSSSVIEKEIFESIKKINEEGKALLEQRTELKHATNPELL
SIYERLLNNKKDRVVVPIENRVCSGCHIVLTPQHENLVRKKDRLIFCEHCSRILYWQESQ VNAQENSTAKRRRRRAAV

11. 11. 526

VFNLEEKSEGVC/UJLKCS_CKENRMESFMIJTDVCQDILGKQKEAVDFFFQAFQPKEAM
OLAEKILGHSGWVFFSGVGKSGCVARKLVATLQSLSERALFFSPVDLLHGDLGLVSPGDI
VCLESKSGETQELLDTVPHLKSRRAILVAITSMPYSNLAALSDLWILDESVAELDFFNL
FINSTTCOMIFGDELAMLEHSRGCVSLSTYGKNHPSGQVGKKANGKVKVDFMFPKTEVPFC
HLGDKVSFSLEVFSAYGCCVCIVDPQFRLMGIFTDGDLRRSLASYGGEVLSLEKVMT
ANPRCITEDSDIAIALQLMESSSPVAVLPVLDNEENRHVTGLLMMHTLAKAGLL

609910 608726

CPR_052/ 609910 606726
sucB-Dihydrolipoamide Succinyltransferase
RYMIFEFRFPKIGETSSGSIVWLKNLGDHVARDEPLIEVSTDKIATELPSPKAGRLVR
FCVNEGDEVASGDVLGLIELEEISEADDESTSCPPTSCETKSEAGSSSSSVWFSPAVLSL AQREGIGIDINIQKIACTGROGRUTRODLEAYISESQQVSIPEIFGGEVNRIPMSPLARAI ASSLSKSSDEVPHASLVVDVDVTDLMNLISGERQRFLDTHGVKLTITSFIVQCLAQTLRQ FPLLNGSLDGTTIVMKKSVNVGVAVNLNKEGVVVPVIHNCQDRGLVSIAKALADLSSRAR LNKLDPSEVQDGSVTVTNFGMTGALIGMPIIRYPEVAILGIGTIQKRVVVRDDDSLAIRK MVYVTLTFDHRVLDGIYGSEFLTSLKNRLESVTMG

611165 609921

CPn_0528 611165 609921
gltT-Glutamate Symport
LMKLWMKIFIGLFVGVTLGLVLEDKAIFFKPIGDIFLNLLSMVVYPLVFCSMVLGIASIS
DMKKLGRIGIKSVGLYLGTTALAIVIGLCFAWIFSFENGCDFAQAQSMBSAVTVIDSNKT
AAYFLSIIAQVFPSNPVRSFAEGNILQIIIFAIFLGIALRLSGERGRPVERFIDGFSEIM
LRMVMMIMSFAPYGVGASMAWISCNHGLGVLWQLGKFIIAYYLACLFHATLVFGGLVRFG
CKMSFSKFLSSMMDAISCAVSTASSSATLPVTMRCVSKNLGVSAEVSGFVLPLGATVNMN
GTAIFQGMAVFIAGAYNCPLSLSSLLLLLVVTATFSAVGSAGVFGGGMITLGSVLASVGL
PIQGIAILAGIDRLRDIVGTPMNILGDAVVATYVASGEGELSPYESIKQESVETT

612298 611165 CPn_0529

CPH_0529 612298 611165
ycaH-ATPase
FSCKEIRAFKRGTMKKRFPSTLFLFYRRVTIAISLEGILGWGWLGSLLSKVFAFLVACWN
RFSWSTPYRARSTVISVGNIVVGGAGKTPTVLWLAEALRLRGYSCGVLSRGYKSQSSRQK
KLITVVDSKVHSASYVGDEPLLMAEKLPEGSVWVHKDRRISAARAAEKFGILLLDGLQYR
KLHKDVEIAVVNGQDPLGGRAFFPKGRLRDFPLRLKTVDAIIVNGGKEAGTVVKRVSNA
PQIFVKPTIASVWTHNGERIPKEALRELRVGVFCGLGFPQGFLNMLREEGIHLLGKYLL
PDHAAITKKELNYFCQOMANRQGQGLLCTEKDSVKLPRLSGEVSLLPIAKVEMRLSVNQD DTLSLLNMIEQIHKNRGN

613323 612460

CPH_0530 51323 512400
spoU-rRNA Methylase
SVVLWGKFLWRRCGSLAFWEFCSMDCIGKHNPLVKEALALKRSRCRKSSWFLVEGAREIQ
KALRTGYLCQHVFCSTHLSEKEKEFLVELKRNSTKLLYCLDSTLAQLSFKEHHDSFVAVL
KRYWNKEDFLIORINAQPFYLIEIEQVEKPGNVGATLRIADGAGVDGVILCNPIVDLYNP
NVRSSLGAVFSLPILSISREEGKELFKQEGWTVFVTSPRAETMYFSKNYLGPTALVFGS
EKDGLTEDWFSEDFSEIALPMLGESDSLNLATSVAAVAYEVVRQRWVN

CPn_0531 614198 613245

CPT_U531 614198 613245
SAM dependent methyltransferase
DSSKDDFRKEKGRRKSQYRDRYVNKDTGRHSKTYFSLIRERLVMDYKLLDSGDGNKLECF
GPVTLIRESIAWPKSRFELMSQAQLQYVREGERGAWKNFKRLPEEMEVAF5DVRCLLK
RTPFGHLGVFPEHMGFWPALKQAIEKHKERQVLNLFAYTGAGSIFAAKGGAVTHVVDASQ
AAVRWAQRNVEKNAFPERRIFWVIEDVISFLKKEIRRNKKYQVILLDPPSYGRGPDGEVF KIDKDLFPLLSLCSKLLADDASYFLLTSHTPGHTPEFLRAIARRSVPTLVSEAWSCGESFCGEGVGALPSGSFVQWIA

CPn_0532 614716 614075
ribC/risA-Riboflavin Synthase
ESFCCKDSVVKWGGMFSGIIQELGEVCFFEAQGNGLSLGIKSTPLFVTPLVTGDSVAVDG
VCLTLTSCNESKIFFDVIPETLACTTLGEKRCSDQVNLEAALKNGDSIGGHLLSGHVFGT AEIFLIKENRYYFRGSKELSQYLFEKGFIAIDSISLTLVSVDSDTFSVGLIPETLQRTTL GKKREGERVNIEIDMSTKIQVDTVKRILASSGKD

614918 615385

CFI_0553
CT406 hypothetical protein
EVAPMQCPFCNHGELKVIDSRNAPEANAIKRRRECLKCSQRFTTFETVELTLQVLKRDGR
YENFQESKLIHGLNAASSHTRIGQDQVHAIASNVKSELLGKQNREISTKEIGELVMKYLK KADMIAYIRFACVYRRFKDVGELMEVLLSATPDMEK

615389 615784

dksa-Dnak Suppressor LNFIRSKVVPLSDDEIEQFKKRLLEMKAKLSHTLEGNAQEVKKPNEATGYSQHQADQGTD TFDRTISLEVTTKEYELLRQINRALEKINESSYGICDVSGEEIPLARLIAIPYATMTVKA QEOFEKGLLSGN

CPn_0535 615763 616296 lspA-Lipoprotein Signal Peptidase KRTPIWKLSSMATRRESTLLVITLFVLIDWYKLVVLLQYKDLQILTHPTLYTHSWGRFS FSIAPVFNEGAAFGLFSNYKYFLFLLRIFVILGLLAYLFFKKKSIQSTTQTALVLLCAGA IGNVGDI I FYGH I VDF I SFNYKQWAFPTFNVADVLI SLGTLLLVYKFYFPTKQTEKKR

010300 617691

CPI_05.6 616300 617601
digA-D_ALL,GIV_Permedse
YRCLGEPLEYFRYTMARCLSLLSVFDDFFWGYVAFILLIVLGVSFSWKSREPGETKESQ
PCKLEPLYGNGATGEREKQSVHPLKVFFASAXSHIGIGIVVGITTAACIGGRGALFWWH
MITCS_LVXYSEVYLAIKERKLDRDGV/QCGPMYFLIKAFKTPVVGVIVAILL/LYGVEI
/GPSVITDSLAHCWALFKVYFMJSLLFLJFYAIFGGLQFIGKICSIVLPFFMLLYCALGL
//IJKESPITLEHULSTVFSSAFFGGSALGGFAGSTTVATTIHQGISRAAYSADIGIGFDSI
IOSESSAKDFTTQAGLITGIA IDALICTHSLDLMVLASGSWISLDLMAAGSVEHTLASYF
PMVKFFLIYEFITVAYTTILSYFLVGKYCAKFLYGNTGAKIYTLYGGLILDEGCFLSQNT
ALLIMOVSGALLIANHULSVFILPKEY/FFFARAASLTETSLSTE

CPn_0537 617315 618189 CT814.1 hypothetical protoin Liellemony/LCSLieccy/LUIGMCTIEVMTICELRQLNKILKNIHRYTTILNFEAKI LAPLMLGKKULGGWUKKRKNRGGLJEDIDELLDEKKQRUWKKNLDQGIKWCAALVLIWKV

618128 618511

CT814 hypothetical protein TKEINGAOHWGSFGKTEVIKIKTIRDIYMFRNNHKPKKTKCKRFRWLRGVLFGGFIATLL TGETT FLANGI PRETIZVRI 2 22 FIN 2201 ING GAKGLARMICHTLOFFK PHILIPPE

618678 CPn_0539 621545

CPn_0539 618678 621545

pmp_19-polymorphic membrane protein

GYNLLGLRHMKQMRLWGFLFLSSFCQVSYLRANDVLLPLSGIHSGEDLELFTLRSSSPTK

TYSLRKDFIVCDFAGNSIHKEGAAFLNLKGDLFFINSTPLAALTFKNIHLGARGAGLFS

ESNVTFKGLHSLVLENNESWGGVLTTSGDLSFINNTSYLCONNISYGPGGALLLGCRKSK

ALFFRDNRGTILFLKNKAVNQDESHPGYGGAVSSISPGSPITFADNQEILFQENEGELGG

AIYNDOGAITFENNFGTTSFSNKASFGGAVVSRYCNLYSGWGDTLFTKNAAKVGGAIH

ADYVHIRDCKGSIVFEENSATAGGIAVNAVCDINAGGPVFINNSALGLNGGAITWQAT

GSILRLHANGGDIEFCCNKVRSQFHSHINSTSNFTNNAITIOGAPREFSLSANEGHRICF

YDPIISATENYNSLYINHORLLEAGGAVIFSGARLSPEHKKENKKTSIINQPVRLCSGV

LSIEGGAILAVRSFYQEGGLLALGPGSKLTTGGKNSEKDKIVITNLGFNLENLDSSDPAE

IRATEKASIEISGVPRVYGHTESFYENHEYASKPYTTSIILSAKKLVTAPSRPEKDIQNL

LIASSEYWGYGYGGSWEFSWSPNDTKEKKTILISWFPTEGEFSLDFKRGSFIPTTLWSTF IRATEKASIELGGVPKVYGHTESFYENNEYASRAYTTSIILSAKLDVTAFSKPENDIQNE IJAESEYMGYGYQGSWEFSMSPNDTKEKKTIIASWTPTGEFSLDPKRGSFIPTTLWSTF SGLNIASNIVNNNYLNNSEVIPLQHLCVFGGPVYQIMEQNPKQSSNNLLVQHAGHNVGAR IPFSFNTILSAALTQLFSSSSQQNVADKSHAQILIGTVSLNKSWQALSLRSSFSYTEDSQ WKKHVFPYKGTSRGSWRNYGWSGSVGMSYAYPKGIRYLKMTPFVDLQYTKLVQNPFVETG YDPRYFSSSEMTNLSLPIGIALEWRFIGSRSSLFLQVSTSYIKDLRRVNPQSSASLVLNH YTWDIQGVPLGKEALNITLNSTIKYKIVTAYMGISSTQREGSNLSANAHAGLSLSF

626862

621631

CPn_0540 626862
pmp_20-polymorphic membrane protein
FiHLIYLSLIEFVNISDRFSSMKWLPATAVFAAVLPALTAFGDPASVEISTSHTGSGDPT
SDAALTGFTQSSTETDGTTYTIVGDITFSTFTNIPVPVVTPDANDSSSNSSKGSSSSGA
TSLIRSSNLHSDFDFTKDSVLDLYHLIFFPSASNTLNPALLSSSSGSSSSSSSSSSSSSSASVAVADPKGGAAFYSNEANGTLTFTTDSGNPGSLTLQNLKMTGDGAATYSKGPLVFTGL
KNLTFTGNESQKSGGAYTEGALTTQAIVEAVTFTGNTSAGQGGAIYVKEATLFNALDSL
KFERNTSGQAGGGIYTESTLTISNITKSIEFISNKASVPAPAPEPTSPAPSSLINSTTID KFERNTSGQAGGIYTESTLTISNITKSIEFISNKASVPAPAPETSPAPSSLINSTTID
TSTLQTRAASATPAVAPVAAVTPTPISTQETAGNGGAIYAKQGISISTFKDLTFKSNSAS
VDATLTVDSSTIGESGGAIFAADSIQIQQCTGTTLFSGNTANKSGGGIYAVQQVTLEDIA
NLÄMTNNTCKGEGGAIYTKKALTINNGAILTTFSGNTSTDNSGAIFAVGGITLSDLVEVR
FSKLÄNTGRIVSAPITKAASNTAPVVSSSTTAASPAVPAAAAPVTNAAKGGALYSTEGLTV
SGLSSILSFENNECONGGGAYVTKTFQCSDSHRLQFTSNKAADEGGGLYCGDDVTLTNL
TGKELFQENSSEKHGGGLSLASGKSLTHTSLESFCLNANTAKENGGANVPENIVLTFTY
TETENEPAPVQQPVYGEALVTGNTATKSGGGIYTKNAAFSNLSSVTFDQNTSSEGGALL
TOKAADKTDCSFTYITNVITNNTATGNGGGIAGGKAHFDRIDNLTVQSNQAKKGGGVYL
EDÄGLLEKVITGSVSQNTATESGGGIYAKDIQLAGSFFIITDNKVETSLTTSTNLYGG EDBITLERVITGSVSONTATESGGGIYAKDIQLQAL PGSFTITDNKVETSLTTSTNLYGG
GISSGGAVPLTNISGTFGITGNSVINTATSQDADIQGGIYATTSLSINQCNTPILFSNN
SAFKKTSTTKQIAGGAIFSAAVTIENNSQPIIFLNNSAKSEATTAATAGNKDSCGGAIA
ANSIELTNNPEITFKGNYAETGGAIGC IDLTNGSPPRKVSIADNGSVLFQDNSALNRGGA
IYGETIDISRTGATFIGNSSKHDGSAICCSTALTLAPNSQLIFENNKVTETTATTKASIN
NLIFAAIYGNNETSDVTISLSAENGSIFFKNNLCTATNKYCSIAGNVKFTAIEASAGKAIS
FYDÄAUNVSTKETNÄQELKLNEKATSTGTILFSGELHENKSYIPOKVTFAHGNLILGKNAE
LSÄVSFTOSPGTTITMGPGSVLSHNSKEAGGIAINNVIIPFSEIVPTKDNATVAPPPLKL
VSRTNADSKDKIDITGTVTLLDPNGNLYQNSYLGEDRDITLFNIDNSASGAVTATNVTLQ
GNEGAKKGYLGTWNLDPNSSGSKIILKWTFDKYLRWPYIPRDNHYYINSIWGAQNSLVTV
KOGSILGNNLNNARFEDDAFNNFWASAIGSFIL KEVLSRUSPTYHGBGYTAAUDAK PROF GNBJARATI I WALDPINSOSKI I LAWFI PRINHFY I PRINHFY IN SIWAGNSLOTV KOGILGINILINIARFEDPAFNIFWASAIGSFLRKEVSRNSDSFTYHGGGYTAAVDAKRRQE FILGAAFSQVFGHAESEYHLDNYKHKGSCHSTQASLYAGNIFYFPAIRSRPILFQGVATY GYMOHDTTTYYPSI EKKMANWDSIAWLFDLRFSVDLKEPQPHSTARLTFYTEAEYTRIR GYKOFLDYDPRSFSACSYGNLAIFTGFSVDGALAWREI ILYNKVSAAVLPVILINIPKA TYBLISTKEKGNVVNVLPTRNAARAEVSSQIYLGSYWTLYGTYTIDASMNILVQMANGGI

CPN_0541 627137 628003
Solute binding protein (-yebi-Synechocystis Adhesin Homolog)
NNRSSYCTAFVMHKVIVFIFTTLYSLKSYGNDVIDKPHVLVSIAPYKFLVEQIAEETCFV
YALUTNHYDPHTYELPPQQIKELRQGDLWFRIGEAFEKTCERNLTCQQVDLSQNVSLQG
KPCCNQHTTNYDTHTMLSPKNLKVQVETIVTTLSKKYPQHATLYQSNGEKLLLALDQLNE
EILTITSKAKQRHILVSHGAFGYFCRDYNFSQHTIEKSSHVEPSPKDVARVFRDIEQYKI
SSVILLEYSGRRSSAMLADRFHMHTVNLDPYAENVLVNLKTIATTFSSL

ABC Transporter ATPase FMTIRILAEGLAFRYGSKGPNIIHDVSFSVYDGDFIGIIGPNGGGKSTLTMLILGLLTPT FGSLKTFPSHSAGKQTHSMIGWVPQHFSYDPCFPISVKDVVLSGRLSQLSWHGKYKKKDF EAVDHALDLVGLSDHHHHCFAHLSGGQIQRVLLARALASYPEILILDEPTTNIDPDNQQR ILSILKKLNRTCTILMVTHDLHHTTNYFNKVFYMNKTLTSLADTSTLTDQFCCHPYKNQE FSCSPH

CPn_0543

628710 629603

(Metal Transport Protein)
KSGIFMLSSLIRDSFPLLILLPTFLAALGASVAGGVMGTYIVVKRIVSISGSISHAILGG IGUTLWIQYKLHLSFFPMYGAIVGAIFLALCIGKIHLKYQEREDSLIAMIWSVGMAIGII FISRLPTFNGELINFLFGNILWVTPSDLYSLGIFDLLVLGIVVLCHTRFLALCFDERYTA LNHCSVQLWYFLLLVLTAITIVMLIYVMGTILMLSMLVLPVAIACRFSYKMTRIMFISVL LN1LCSFSGICIAYCLDFPVGPTISLLMGLGYTASLCVKKRYNPSTPSPVSPEINTNV

630548 029525

CPIL 0544 - 0.3548 - 0.3525

YINDZ-GTP binding profein
KRBYFCES LKSSFFCLKKDKNV LMFVDQ LTLELRAGKGONGVVAWRKEKYLPYGGPYGGN
GRNYSEV LLEATTGVV SFEAYRN LFFLKAPEGO GAMINRFORSOKDLIVSVPFOTLLRD
ARTGELLHDETVOGERLVGGGGKYSKGNTFFKTSVIRAPTKATFGKPGELRGVELELKL
FAD LGLVSFFNA SKOTLMTLAHTEVKVAAYPFTTLAPSLGLVLCKDRLYGKWITADIP
SH LFFABRONKGD SLOTLENTLAHTEVKVAAYPFTTLAPSLGLVLCKDRLYGKWITADIP
SH LFFABRONKGD SLOTLENTLAHTEVKVAAYPFTTLAPSLGLVLCKDRLYGKWITADIP KDMUVALNK IDDELLIDEQLEC LQSFQKRE PSYTEVL ISOUTGEGVIX LYRFFTQRLAV

CTH_0545 630908 630633 rls7 ls2 ribesonal protein TEAMEERMAHKKKKISHMERDEKSKRLJVKVGAROKVTTGJILVRORGTPMNPAQAVQ TUTTUR COMMUNICAN FORTSTIL

630912

CPn_0546 = 51243 = 530312 F121-621 Ribosomal Protein ESKORLTUSIERFIRKKLMEPYAVIOTOSKQYQVRGODVIDVELLGE/AGDKEVIFQDVL FVFDGTKASLGSPTIANAQVKAEYLSHVKGEKVVAYKYKKRKNYHRKHGHRQKYLRVKIR

631589 632188 CPn_0547

FGCGDGVQCFCVLTVMEYCD

633234 632191

CPH_U548
cysJ-Sulfite Reductase
kmyLoekfkaQQvPLvLreLLsCsdsIndsDpiyrmvfdsndTTiSykVgDaLgvLpens
kevsehvLQLLgySpTTLvwvkKTsEkvsaQkFiQGvVDLDKIPaKLnSffPDkDPkITL
yDaIQeyrpQIPieLfaEsVFPLLPRFYSIASSPDLHPKSIELLvkHvSyPGkyQkRfGv
CSSFLCSELQvxDsAYIFVQPTKHFTLSTQTEGKPLVMIGAGTGIAPYXAFLEERLFNKD
FGNnLLffGERKEKVNFYYREFWNHAEEEGKLKLFLAFSRERDQKVYVQDLLRIQKDEVR

KAYEEGGFFFVCGRKVLGIEVKHALEEILGKDTLASLRKEHRYVVDVY

CPn_0549 633662 rs10-S10 Ribosomal Protein 633255

PODVOHOPWNOHSLLRFLKKFKKLLRSKGCMKQQKQKIRIRLKGFDQGQLDRSTADIVE TAKKTGARVVGPIPLPTKREVYTVLRSPHVDKKSREQFEIRTHKRLVDILDPTGKTIDAL KMLALPAGVDIKIKAA

635688 633580

fusA-Elongation Factor G LIYSENIKFMSNQEFDLSAIRNIGIMAHIDAGKTTTTERILFYAGRTHKIGEVHEGGATM DWMAQEQERGITITSAATTVFWLGAKINIIDTFGHVDFTIEVERSLRVLDGAVAVFDAVS GVEPQSETVWRQADKYGVPRIAFVNKWDRWGADYFAAVESWKEKLGANAFPVHCPIGSES QFVGMVDLISOKALFFLDDTLGAKWEEKEISEDLKERCAELRANLLEELATIDESNEAFM GFVGMVDLISQKALYFLDDTLGAKWEEKEISEDLKERCAELRANILEELATIDESNEAFM
MKVLEDPDSITEDEIHQVMRKGVIENKINPVLCGTAFKNKGVQQLLNVIVKWLPSPLDRG
MIRGINLKTDQEISLEPRRGGPLAALAFKIMTDPYVGRITFIRIYSGTLKKGSAILNSTK
DKKERISRLLEMHANERTDRDEFTVGDIGACVGLKFSVTGDTLCDDNQEIVLERIEFPDP
VIDMAIEPKSKGDREKLAQALSSLSEEDPTFRVSTNEETGQTIISGKGELHLDILRDRMI
REFKVEANVGKPQVSYKETITVSGNSETKYVKQSGGRGQYAHVCLEIEPNEPGKGKEVVS
KIVGGVIPKEYIPAVIKGIEEGLMTGVLAGYGLVDVKVSIVFGSYHEVDSSEMAFKICGS
MAVKDACRKAKEVILEPIMKVAVITPEDHLGDVIGGLMRRGKILGQESSRGMAQVNAEV PLSEMFGYTTSLRSLTSGRATSTMEPAFFAKVPQKIQEEIVKK

CPn_0551 636174 635698
rs7-s7 Ribosomal Protein
MYMSRRHSAEKRDIPGDPIYGSVILEKFINKVMMHGKKSVARKIVYSALERFGKKLNLEN
VLEGFGEALENAKPILEVRSRRVGGATTQVPVEVASERRNCLAMQWIIKHARSKPGKSME
VGLATELIDCFNKQGATIKKREDTHRMAEANKAFAHYKW

636219 636698

rs12-S12 Ribosomal Protein
IQAGYVPSSSENKPLPTKRALLYISMLVVVRLKREEYMPTINQLIRKRRKSSLARKKSPA LOKCPOKRGVCLOVKTKTPKKPNSALRKVAWVRLSNGQEVIAYIGGEGHNLQEHSIVLIQ GGRVKDLPGVRYHIVRGTLDCAAVKNRKQSRSRYGAKRPK

637753 636812 CPn_0553

CPn_0553 636812

No robust homolog present in Genebank/EMBL as of 11/7/98

GCMWRVVLRFLIIFILGRAVFPLRASESFSWETSTCLTVLGIPFIDIILTTNEDFVAQCG
LQIGTISSTNNAKIKEIFLIYKEKFPEASISFKRKEPLNLSQSHLSDLGILGWRNGETYA
EGMANKENGPALKQPKOLRLVLRCCPNQPOTLLYSEKEAEKGIETNTCLCNQGYTLLDGQL
ILYGDSIEKFLKETKRKNNHTLVDLCDSQVVTTFLGRFWSLLNYVQVLFLSEDSAKILAG IPDLAQATQLLSHTVPLLFIYTNDSIHIIEQGKESSFTYNQDLTEPILGFLFGYINRGSM EYCFNCAQSSLGET

CPn_0554 637806 638141
CT440 hypothetical protein
VFSVLLLCIILVYVRFMYECKSRMASPTPQQLHLQQKVESKAYDYSRSLAMIATALLFFI
VALILSGLSLLPQVFLPFSGAYFIIGSFLAFIALGILLINCVCDLKQYLTSS

CPn_0555 638298 640241
tsp-Tail-Specific Protease
MFVMKKLVRLCVVLLSLLPMVLFSSDLLREEGIKKMMDKLIEYHVDAQEVSTDILSRSLS
SYIQSFDPHKSYLSNGEVAVFLQSPETKKRLLKNYKAGNFAIYRNINQLIHESILRARQW
RNEWVKNPKELVLEASSYQISKOPMOWSKSLDEVKQRQRALLLSYLSLHLAGASSSRYEG
KEEQLAALCLRQIENHENVYLGINDHGVAMDRDEEAYQFHIRVVKALAHSLDAHTAYFSK KEEQLAALCLRQIENHENVYLGINDHGVAMDRDEEAYQFHIRVVKALAHSLDAHTAYFSK
DEALAMRIQLEKGMCGIGVVLKEDIDGVVVREIIPGGPAAKSGDLQLGDIIYRVDGKDIE
HLSFRGVLDCLRGGHGSTVVLDIHRGESDHTIALREKILLEDRRVDVSYEPYGDGVIGK
VTLHSFYEGENCYSSEQDLRRAIQGLKEKNLLGLVLDIRENTGGFLSQAIKVSGLFMTNG
VVVVSRYADGTMKCYRTVSPKKFYDGPLAILVSKSSASAAEIVAOTLQDYGVALVVGDEY
TYGKGTIGHGTITGDASQDDCFKVTVGKYYSPSGKSTOLQGVKSDILIPSLYAEDRLGER
FLEHPLPADCCDNVLHDPLTDLDTQTRPWFQKYYLPNLQKQETLWREMLPQLTKNSEQRL
SENSNFQAFLSQIKSSEKTDLSYGSNDLQLEESINILKDMILLQQCRK

640921 640325

CPR_U55% 040921 040325 crpA-15kDa Cysteine-Rich Protein ENGMSSNLHPVOGTGTGAAAPESVLNIVEEIAASGSVTAGLQAITSSTGMVNLLIGWAKT KF TQPIPESKLPQSRACQITLLVUGILLVVAGLACMTIFHSQLGANAFWLIIPAAIGLIK LLVTSLGTDEALTSEKLMYFOKWAGVLEDQLDDGILNNSNKIFGHVKTEGNTSRATTPVL NDGRGTPYL3PLVSKIARV

612870 641194

CPE_0557 6.2870 6.41194

omcB + 0kDa - Cytterine-Rich OMP

EIPMEKLIRRNTPULNEPMAGGRAGOFIEAAVARGLITKIVAGAETKRARVEMTARKVR

EIPMEKLIRRNTPULNEPMAGGRAGOFIEAAVARGLITKIVAGAETKRARVEMTARKVR

EVRRIKQPVEQA, RIGARCIDKEF*PCETGRCQP*PCAQQEGCYGRLYSVKVNJJK;NVETQGS

VPEYATVGGTY-FILLILATERKD* 7DVVTTOQL;PCEABFVGGDPETTTP*IDTKLWR FIDRI,

GAGDKCK ETWWKPILKERXC*PTAATVCAC PELRGYTRKGODPA LGTKQTBKXOCHKLOPVC

YK LEVVITYLYLARNTTVONPVPEXTYGHACQQRVLGPNLGDMRPGDKKVFTVET**PQPRG

QTTNVAT*TTYGGBHK_JANGHTY-TITET** VQVNLGTJADMGYVCKPVEY-1CGV.MINTODYLH

LVV LQDTJPNSTYTULA BANGHTC FINYAMP LEPMETGGETGDFKLVVKAQVEGBHTMGVAV

TGESNCGTCTSCAETTTHWKGLAATHMCVLJYNDPICVGENTVYRTCVTNRGJAEDTHM;

LILKETKELQPIA TERPTKGTT ENTTYPFDALPKLAGKEGVEFGVTTKGTAEGDAAGGAT

GSSDYGTSHVSDTENTHVY

643305

omcA-9kba-cysteine-Righ Lipoptotein KLMKKAVLIAAMFCGVVGLSSCCRIVDCCFEDPCAPSSCNPCEVIRKKERSCGGNACGSY VPSCSNPCGSTECNSGSPQVKGCTSPDGRCKQ

+43700 643937

TOTALLE TREPORTOR TO THE TOTAL CONTROL OF THE OPPREYE HEAVER FOR SANKAHULD FFLARSVENTCYNTNL

CPn_0560 645666 644098
gltX-Clutamyl-tRNA Synthetase
RNSRFQGMKSLWSKDKRIMWNENVRVRVAPSPTGDPHVGTAYMALFNEIFAKRFKGKMIL
RIEDTDRTRSRQDYESNIFSALRWGIQWDEGFDVGGPYGPYRGSERTKIYQGVVETLLK
TDCAYKCFATPQELAEMRAVASTLGYRGGYDRRYRYLSPEEVASREAAGQPYTIRLKVPL TICLAYKCY RIYOGLAEMRAWS; ILSY KOGYDRKYN ISPEEVAKRAAGOPY I IRLAVPE SGECVFEDYSKGRVVF PWADVDDQVLVKSDGF PTYHFANVIDDHLMGITHVLRGEEMLSS TPKHLLLYEAFGWEPPVFLHMPLLLNPDGTKLSKRKNPTSIFYYRDSGYVKEAFVNFLTL MGYSMEGDEEVYSLERI I ETFNPRRI IGKSGAVFDIQKLDWANKHYLNHEGSPECLLKELQ GWLLNDEFFLKILPLCQSRITTLAEFINLTSFFFSGLLEYRVEELLPQALSPEKAAILLY SYVKYLEKTDQWTKETCYLGSKWLAQAFNVHHKKAI I PLLTVAITGKKQGLPLFDSIEIL GKPRARARLVYAEKLLGGVPKKLAATVDKFMQREDFSEATFDL

646407 645871

euo-CHLPS Euo Protein

LMACEQJEGCYELEEREEIEDIKDSDTKWVSITQAAKLHNVTRQAIYVAIKQKKLKASKE TRWEIDIKDLEEYKRNRYSRKKSLYQGELVFDNGKGCYSINQVAQILGIPVQKVYYATRT GTIRGERKGAAWVIHVSEIERYKNEYLSKQAAKKLKGAEPKEHQAPNFEPPTEIFPESN

CPn_0562 648051 6469 *CHLPS 43 kDa protein homolog_1 646918

TCHEPS 43 ALD DICTERIN HOMOGOTI WAS ALD DICTERING WAS ALD DICTERIAL SDDVIKEVAGHIQRHFDN

650113 648293

recg.ssdna Exonuclease QYMMLLWDFSPKGPCGIKFMTNSDNASAAGLLWAHPKEDPAFLGMIIKEFHLPPTVAQIF QYKALLWDFSPKGPCGIKFMTNSDNASAAGLLWAHPKEDPAFLGMIKEFHLPPTVAQIF
ISRGFQTIQEIHKFLYSHLSSLYDPGLFLDMSKAVERLLLARDRKEHVMIYGDSDVDGMT
GVÄLLVEFLRDIDVHVSYFFLGAILKQHGETSTLIAKLKEBGITLLITVDCGITAGKEVS
DIFFRGGIDVIITDHHMPTGKIPHCVATLNPKLRDHTYPNRELTGVGVAFKLARGVLNALI
SRNSVPKSQGSLKKLLDLVTLGTITDVGVLLGENRVMVRYGIKEIARGARPGLNKLCALC
GVEKSEVTSTDIVLKIAPKLNSLGRLDDPAKGVELLLTQDDERVDALHELDNINRERQR
IEAEVFQDVQEILNSNPEILKQAAIVLSSTAWHARVIPIISARLAKTYNKPVVIIAIQRG
IGKĞSARTIGSFPLLGVLKKCSSLLLSYGGHDFAAGVIMKEDKVEDFKKKFVHLVNSSLK KGPTLPHLEIDAYADFDAIDYDLLASMELFEPFGKGNLMPIFYSKVRQVRYPKVLPGNHL KLYLSQKERNLEGVAFGLGRHADALKASWHYPLEIAYTPRLSQTSGSGVIHLLVRDFRIS SEPRESD

654359 650145

CPB_0564 654359 650145

sedDksecF-Protein Export Proteins SecD/SecF (fusion)

SGAMKOKVKRNFAIICVFALALYYVLPTCLYYAKPLDKKIDGNEAEHIIKSFTKQAQQV

KKBVTPRVSAILSSLHLRGHIQQHPAIPDIVSVRFKRGEDAEDFIGNLVHGEPNVPIKSA

RLHUNGYSREHDDHVIQVASSINTSLVESDFSFVSYSSENEQEWASSILQRYYSACTFPK

KKBVTPRVSAILSSLHLRGHIQQHPAIPDIVSVRFKRGEDAEDFIGNLVHGEPNVPIKSA

RLHUNGYSREHDDHVIQVASSINTSLVESDFSFVSYSSENEQEWASSILQRYYSACTFPK

KKELTLHSDLLAQRTSLSKEQRLDFDSRLAVEKQKLSKNLTVQVEDYNNGFSFQWMDKD

TGKÄILQCERLLGGIAEHLTAILTHRPAAESCOLIPENFVFCRQPBESAFGCYIFSP

NTDCHFSKGSVYILLKGLRSIVAKYQQGGKELQSFEKDLQNLYNCFSHTEAISWTLGE

DQVLLIRRPLQOFLDVWGEGFVIGKEGCAFLEVKDIQDRLATVNQIEKNRGSDLVRWHEQ

VRHAKCSMDLQERLSAPIPYONLFLENNKLINWKFSRGENILRIGIDFVGGRQLLISFKD

HOGKGLTDKEDILKVSDELCARLNKLGVSEIELRREGDYIHLSVFGSSTISSEILGTSK

MSFHVVNERFSSYSASRYEVQRFLDYLWFTSQAQGKTSPEEINTFASALFNEEVDVPPSK

HEAITKLKSEGLAFSPSGCETPSTDLDTTFSMIAIGKDAEQKANPLVIVFRNYALDGASL

KDIRREFAAGEGYVLNFSVKDTSFKKMAEKLSPTESFHTWTSAVCQEGISGTANGQYSAN

RGWRMAVVIDGYMVSSPILNVPLKNHASVSGKFTHREVSKLASDLKSGAMSFVPEVLSEE

TISSDLGKKQCTQGIISACCGLAMLIVLMSVYYRFGGVIASGAVLINLLLUMALQYLDA RGWRNAVVIDGYMVSSPILNVPLKNHASVSGKFTHREVSKLASDLKSGAMSFYPEVLSEE
TISSDLGKKQCTQGIISACCGLAMLIVLMSVYYRFGGVIASGAVLLNLLIWAALQYLDA
PLTLSGLAGIVLAMGMAVDANVLVFERIREEFLLSQSLKKSVEKGYTKAFGAIFDSNLTT
VLASALLFFLDTGPIKGFALTLILGIFSSMFTALFMTKFFFMLWMNKTQHTQLHMMNKFV
GIKHDFLRGCKKLWAVSGSVFLLGCVALGFGAWNSVLGMDFKGGYAFTFNPKEHGISDVA
QMRGKVVHKLQEAGLSSRDFRIQTFGSSEKIKIYFSDKALSYTKADTSLSPKINDHELAL
AVGLLSETGLDFSTETLNETONFWSKVSSKLSKKMRYQATIGLLGALAIILLYVSLRFEW
OYAFSAVCALHHOLATCAVLFIAHFFLKKIQIDLQAIGALMTVLGYSLNNTLIIFDRIR
EDRQANLFTPMHVLVNDALQKTFSRTVMTTATTLSVLLMLLFIGGSSVFNFAFIMTIGIL
LGTLSSLYIAPPLLLFMVRKENRSK

CPn_0565 655741 654533
CT449 hypothetical protein
NKLECFLIFCFVNISAILFDSSFLLKIKRNSKRMLRSMKFPRISISDLIPTOMVIWWRGG
GNYHYVPNAQNLPKKILGGVLACFGLALLGCAAFAAGVCQTIFPCIGLMILGLVLLGFAY
LQYSKGWSRFERPLFRETKVFEKPINWLGCLSLLQSWKKIRPGCYYHPGCPQVEICEGSQ
EIVTKIFOKKSDRNTSIFLIGEMDQIALRQGIEKSSLSRKTFAIDPSVVSSLLSEIQREE
QYLLDFKVISWSSEDDASDRTHPKSAIYVNISDAAGEPGRCYIDAYTKAFFTVLDQIGD
ENIVKKITIYVLTPILGVPDALPKEEQENLKLLSQAAFLYSAEQVAKRMREEKQDSIRIK
FIETDPTSPTGLYFSPHHGSTPHSVTPISLSGFVGEQESYTFA

CPn_0566 656099

yood tamily

A HOLDER FOR THE PROPERTY OF T

656894

odoA Phospharidate Cyridylytransterace

VLNSNKFKSKTGAYGDLFGPVVVHSLVUTFLVLLLYGSLFFLTGFALGFITATCGAVGTY
EYSSMAKAKMHYPLSTFSAIGSFLFLALSFLSIRWGHGLBOFFDALPWTLLIVWVVWSIF
RVRKSTIGALOLSGVTLFSILYYGIPIRLFLHVLYGFIHTQEPYLGIWAGFLIATTKGA
DIFGYFFGKAFGNKKIAPQISPNKTVVGFVAGGLGATLIGFIFFLQIPTRFASYFFMPAI
LIPJGLALGITGFFGDIIESIFKRDAHLKNSNKLKAVOGMLDTLDSLLLSTPIAYLFLLI TOSKEFIG

CPn_056R 557805 658464

EARLY FOR EARLY FOR THE CONTROL OF THE CON

658398

659044 660789

args-Arginyl trna transferase

TKLESSKHGMNRGAKETSPKLMSTLLSLSVICSQAIAKAFPNLEDWAPEITPSTKEHFG

HYQCNDAWLARVLKKAPRAIAEAIVAELPGEPFSLIEIACAGFINFTFSPVFLNQOLEH

FKDALKLGFQVSQPKKIIIDFSSPNIAKDMHVGHLRSTIIGDSLARIFSYVGHDVLRLNH FKDALKLGFQVSQPKKIIIDFSSRNIAKDMHVGHLRSTIIGDSLARIFSYVGHDVLRLNH
IGDWGTAFGMLITYLQENPCDYSDLEDLTSLYKKAYVCFTNDEEFKKRSQQNVVALQAKD
PQAIAIWEKICETSEKAFQKIYDILDIVVEKRGESFYNPFLPEIIEDLEKKGLITVSNDA
KCVFHEAFSIPFMVQKSDGGYNYATTDLAAMRYRIEEDHADKIIIVTDLGQSLHFQLLEA
TAIAAGYLQPGIFSHVGFGLVLDPQGKKLKTRSGENVKLRELLDTAIEKAEEALREHRP
LTDEAIQERAPVIGINAIKYSDLSSHRTSDYVFSFEKMLRFEGNTAMFLLYAYVRIQGIK
RRLGISQLSLEGPPEIQEPAEELLALTLLRFPEALESTIKELCPHFLTDYLYNLTHKFNG
FFRDSHIQDSPYAKSRLFLCALAEQVLATGMHLLGLKTLERL

CPn_0571 662179 660749
murA-UDP-N-Acetylglucosamine Transferase
TFEKVNV5FSDFDAKGERRMQIAQVFGCGRLNGEVKVSGAKNAATKLLVASLLSDQKCTL
RNVPDIGDVSLTVELCKSLGAHVSWDKETEVLEIYTPEIQCTRVPPTFSNVNRIPILLLG
ALLGRCPEGVYVPTVGGDAIGERTLNFHFEGLKQLGVQISSDSSGYYAKAPRGLKGNYIH LPYPSVGATENLILAAIHAKGRTVIKNVALEAEILDLVLFLQKAGADITTDNDRTIDIFG
TGGLGSVDHTILIPDNIEAASFGMAAVVSGGRVFVRNAKQELLIPFLKMLRSIGGGFLVSE
SGIEFFQERPLVGGVVLETDVHPGFLTDWQDFFAVLLSQAGGSSVIHETVHERNLGYLHG
LQHMGAECQLFHQCLSTKACRYAIGNFPHSAVIHGATPLWASHLVIPDLRAGFAYVMAAL IÄEGGGSIÏENTHLLDRGYTNWVGKLRSLGAKIQIFDMEQEELTTSPKSLALRDASL

664616 662349

CT456 hypothetical protein
IMAAPINOPSTTTQITQTGQTTTTTTVGSLGEHSVTTTGSGAAAQTSQTVTLIADHEMQE IASODGSAVSFSAEHSFSTLPPETGSVGATAQSAQSAGLFSLSGRTQRRDSEISSSSDGS SISRTSSNASSGETSRAESSPDLGDLDSLSGSERAEGAEGPEGPGGLPESTIPHYDPTDK ASILINFLKNPAVQQKMQTKGGHFVYVDEARSSFIFVRMGWSTAESIKVSNAKTKENITK
PADLEMCIAKFCUGYETIHSDWTGRVKPTMEERSGATGNYMHLMLSMKFKTAVVYGPMNA
KESSGYTPSAWRGAKVETGPIWDDVGGLKGINWKTTPADDFSFINETPGGGAHSTSHT
GPGTPVGATVVPNVNNLGGIKVDLGGINLGGITTNVTTEEGGGTNITSTKSTSTDDKVS
ITSTGSQSTIEEDTIQFDDPGQGEDDNAIPGTNTPPPPGPPPNLSSSRLLTISNASLNOV ITSTGSGSTTEEDTTQFDDGGGEDDNATGSTNTPPFGPPFDFNSSSRLDTISNASLNQV LQNVRQHLNTAYDSNGNSVSDLNQDLGQVVKNSENGVNFPTVILPKTTGDTDPSGQATGG VTEGGHIRNIIORNTOSTGQSEGATPTPOPTIAKIVTSLRKANVSSSSVLPOPQVATTI TPQARTASTSTTSIGTGTESTSTTSTGTGTGSVSTQSTGVGTPTTTTSTGTSATTTTSS ASTQTPQAPLPSGTRHVATISLVRNAAGRSIVLQQGGRSQSFPIPPSGTGTONMGAQLWA AASQVASTLGQVVNQAATAGSQPSSRRSSPTSPRRK

CPn 0573 665413 664691

yebC family VEDMAGHSKWANTKHRKERADHKKGKIFSRIIKELISAVKLGGADPKSNARLRMVIQKAK ENNIPNENIERNLKKATSAEOKNFEEVTYELYGHGGVGIIVEAMTDNKNRTASDMRIAIN KRGGSLVEPGSVLYNFARKGACTVAKSSIDEEVIFSYAIEAGAEDLDTEDEENFLVICAP SELASVKEKLISQGATCSEDRLIYLPLRLVDCDEKDGEANLALIDWLEQIEDVDDVYHNM

CPn_0574 665978 665394
No robust homolog present in Genebank/EMBL as of 11/7/98
SAERGFRHPIVMVETVLHNFQRYLSKYLYRVFRFPCRKKTFLSSHRVLARPSFPVDYCPG
KIYDLQEIYEELNAQLFQGALRLQIGWFGRKATRKGKSVVLGLFHENEQLIRIHRSLDRQ EIPRFFMEYLVYHEMVHSVVPREYSLSGRSIFHGKKFKEYEQRFPLYDRAVAWEKANAYL LRGYKKRVGGGYGRA

CPn_0575 eco524 665982
YhhY-Amino Group Acetyl Transferase
SIFGRVWRSFMTAEKONTGILGLEIRYTLPSDATYMLKWLNDPKILRGFPIQTEAEIRET
VNFWVGFYRYHSSLTAVYNGNVAGVATLVLNPYVKVSHHALISIIVGEEFRNKGIGTALL NNLIHLAKTRFKLEVLYLEVYEGNPALHLYQRFGFVEVGRQNRFYKDEIGYLAKTTMEKD

CPn_0576 6e7543 666494 prfB-Peptide Chain Release Factor 2 (natural UGA trame-shift) MQENLDKRLEALRTEISLAARSL

CPn_0576.1 667548 prfB-(natural UGA trame MQENLDKRLEALRTEISLAARCL trame-shift)

CPn_0577 CANTS & REPARE, 058155
DWIB (YM74) COMPLEX PROFESS
ECHMOQKNKNSAFMHDVNTSTDLAVIVOKOFMPRTEIVKKVWEYTKKHNCOPOKNKRNIL
PDANLAKVEGSSDEIDMECMTRALGKHIVK

CPn_0578 668388 008338 6,6,036,5

CPE_UT/A POSTO DESIGNATION OF THE PROPERTY OF

LKLLKNTPUTLLHNTTHVIPNTLNIVGLGDLFARQFHPEQAFKNYDPSLPGLLLSHNPDG ITRLQQYPGDFVLSGHSHGPQVTLSWPKFARKFFERLSGLENPYLARGYFVTKEGKQLYV NRGL&GLKRIRFCSPPEICYITCSYD

CPn_0579 669310 669993
ygbP/yacM-Gugar Nucleotide Phosphorylase
KEFASAPLLKGATGHVPMIKSSLILLSGOGGTRFGSKIPKQYLPLNGTPLVLHSLKILSS
LPOIAEVIVVODPSYOETFQEYPVSFAIPGERRQDSVPSGLQQVGYPWYIIHDGARPFIY
GEGHERLEGTARE FOR TALMAFIL TEMPORPVOT ERDBLAIFFT I KTGILPFGLA
LAKEKQUTUVDD LBARELDFE GUVTNKHDQIKI GPUDDTIAGAGU

669936 670793

CPH_U580
truA-Pseudouridylate Synthase I
ASSNONFLPRRSNDCPSPPMTKVALLIAYQGTAYSGWQQPNDLSIQEVIESSLKKITKT
RTPLIASGRTDAGVHAYGQVAHFRAPDHPLFANANLTKKALNAILPKDIVIRDVALFDDN
FHARYLAIAKERYRSISRLAKPLPWQRHFCYTPRHPFSTELMQEGANLIGTHPFASFAN
HGRDYNSTVRTIYTLDIVDKGDSLSIICRGNGFLYKMVRNLVCALLDVGKGAYPPEHLLD ILEQKNRREGPSAAPAYGLSLHHVCYSSPYNNFCCEQCSVSTSNEG

671533 670745

CPH_0581
Phosphoglycolate Phosphatase
EGLRWRSVKSFLRQCWIYSMLVSDEFQLCLRSGMYLEDYDVFFFDLDGLLVDTEPCFYRA
FLQACAEFSLEVHWDFSTYYSHTTLGTEIFSKKFIEQYPQAQEYMAEIFAKRLQIYYKSL
EHACPALMPGVEAFIELVLSLNKTFGVVTNSPRDATHTLRTMYPILNKFLFWVTREMYAR
PKPYGDSYDYAYRTFAREGMKVIGFEDSVKGLRALSKIPATLVCINSMAEITPEDYPELK
CMEDEGYDBEDDULTBURGCOMUL GKEFFSYPSFDVLTEHCSOOKLL

671305 672177

C71303
C7465 hypothetical protein
KNPNALLKKIOHRLVKMHDKNKVLYLQANHLNQKRKHHPLNTYHSSNTTETRRLPTYYK
SNIVLKMILRISTVSLLTSCSFSKNSRTCFVTPERITSQKDCPVLLHPKSTTISPPLYDW
ISPNREVITAYSFYCRGQGNSIITPEGVLYDCDGLHHSITKEEFRYIHPRLIEVVRLLQQ DHPKVSIIEAFCCPKHFHFLEASGISLSQLHLQGTAATFALDPPLPMEKLLATIKKLYKK NSDPSLSNFIVTEATLTNPELRLTQQDLGSHTEITVEILDNLQNKEALSSA

672349

CT466 hypothetical protein
IVLSFFLGKTKVTPRFLMNERTLLLLLKKKKGLFLAILDLTQTESSLTTPELEKVLKQKK
IFLSCIDRVDLQIKEFRHAFSSELPQDIQEELEEIRDVIIRILDTDKRNYAQKKKEFGIY

CPn_0584 672659 673798
ato\$/ntrb-2-Component Sensor
IRTMATMHRKKRNLVFMNVPDSKNLHPPAYELLEIKARITQSYKEASAILTAIPDGILLL
SETGHFLICNSQAREILGIDENLEILMRSFTDVL.PDTCLGFSIQEALESLKVPKTLRLSL
CKESKEKEVELFIRKNEISGYLFIQIRDRSDYKQLENAIERYKNIAELGKMTATLAHEIR
NEASGIVGFASILKKEISSPRHQRMLSSIISGTRSLNNLVSSMLEYTKSQPLNLKIINLQ
DEESSLIPLLSVSFPNCKFVREGAQPLFRSIDPDRMNSVVMNLVKNAVETGNSFITLTLH
dddiscumpott Deet Manye Tenepayerde As A ON LIBIT MECOLOGY MEDERATMED AND LIBITATURE DEAD VILLE MECOLOGY MEDERATMED AND LIBITATURE DESCRIPTION OF THE PROMETER OF THE PR TŠGŠĪSVTNPGTIPSEIMDKLFTPFFTTKREGNGLGLAEAQKIIRLHGGDIQLKTSDSAV SEFIIIPELLAALPKERAAS

CEC0585 675880 673865
*SIMIlarity to Cps IncA_2
ISERKILRPINFSIGECSSMMATPAQKSPTFQDPSFVRELGSNHPVFSPLTLEERGEMA
IARVOCCOWNHTIVKVSLIILALLTILGGLLVGLLPAVPMFIGTGLIALGAVIFALALI IARVQQCGWNHTIVKVSLITLALLTILGGGLLVGLLPAVPMFIGTGLIALGAVIFALALI
LGLYDSQGLPEELPPVPEPQQIQIEDLRNETREVLEGTLLEVLLKDRDAKDPAVPQVVVD
CEKRIGMLDRKLRREEEILYRSTAHLKDEERYEFLLELLEWRSLVADRLEFNRRSYERFV
QGIMTVRSEBGEKEISRLQDLISLQQQTVQDLRSRIDDEQKRCWTALQRINQSOKDIQRA
HDREASQRACEGTEMDCAERQOLEKDLRROLKSMQEWIEMRGTIHQQEKAWRKONAKLER
LQEBLRIGIAFDEQSLFYREYKEKYLSQKLDMQKILQEVWAEKSEKACLESLVHDYEKQ
LEQKDANLKKAAAVWEEELGKQQQEDYEQTQEIRRLSTFILEYQDSLREAEKVEKDFQEL
QQRTSRLQEEKQVKEKILEESMNHFADLFEKAQKEMMAYKKKLADLBGAAAPTEIGEDDD
WYLDDSASLSQKKIRELVEENQELLKALAFKSNELTQLVADDAVEAEKEISKLREH IEGOK
EGLRALDKMHAQAIKDCEAAQRKCCDLESLLSPVREDAGMRFELEVELQRLQEENAQLRA
EVERLEQEQFQG

TEMPOSE 67593 677183
atoC/ntrc-2-Component Regulator
KEKNNPSRGENMAIKNILVVDDEPLIRENFLSELLTSQGFIPDTAENLRNALOMIRSRDYD
KUTSDMSNPDGSGLDLIKIIKOSSPHTPVLVVTAYGSIENAVEAMHQGAFNYLTKPFSSE
ALFAFISKAEELKNLVHENLFLHSQTTPDSHPLIAESKAMKDLLAIAKKAASSSANIFIH
GESGCGKEVLSFFIHHNSPRANHPYIKVNCAAIPETLLESELFGHEKGAFTGATTKKAGR
FELAHKGTULLDEITEVPVNLQAKLLRAIQEKEIEHLGGTKTLSVDVRILATSNRKLKEA
LDXBSFRODLYYRLNVIPLHLPPLRDRODDILPLANVFLNKFCRMNNTPLKTLSPKAQEL LLNYPWPGNIRELSNVLERVVILENTSLLTEDMLALA

CPn_0587 677378 678124
*yvyD_Bs conserved hypothetical protein
syceLfILsTLLKHHVTLGDKMRPHRKHVSSKSLALKQSASTHVEITTKAFRLSMPLKQL
ILEKSDHLPPMETIRVVLTSHKDKLGTEVHVVASHGKEILOTKVHNANPYTAVINAFKKI
RTMANKHSNKRKDRTKHDLGLAAKEERIAIOEEQEDRLSNEWLPVEGLDAWDSLKTLGYV PASAKKKISKKKMSIRMLSQDEAIRQLESAAENFLIFLNEQEHKIQCIYKKHDGNYVLIE

CPn_0588 678033 678626
CT469 hypothetical protein
TSKSIKSNAFIKNMTATMSLLNLPSSQDSACEDSTSQSQIFDPIRNRELVSTPEEKVRQR
LU3FLMHKLNYPKKLIIIEKELKTLFPLLMPKOTLIPKRRPDILIITPFTYTDAQGNTHN
LGDPKPLLLIECKALAVNQNALKQLLSYNYSICATCIAMAGKHSQVSALFNPKTQTLDFY PGLPEYSQLLNYFISLNL

678634 679395

CMILOSMY AND THE CONTROL OF A C ONTO TEKEBBOTION THAKOFSELLA TABEPTATABETTYLFD: LOBERKSERNOSEDP

CPn_0590 680136 679516 CP471 hypothetical profess

LFLYGDHNLGFACPYLFFFTVLFASGSFGNQLLSVPTWLGEBEJFYTHPFDFDKSYPDME NMETQAORKKRVEFNLTGEPPKLSTLNYQGSFGHLRAKGRGVYPVLYALNFGCSSCKMDM DFRGKWNRSSTITISNQKESINLKLPKDVGVIVNTKTGLKGNVCPGSTFIKQGWGVWNKI YHNDLVGFSEVTLIFNVSSEGGTITFS

681020 680364

VagE Eamly
SLIMRCTAYCTASAYNLHVLFHLLKPRYFTILSREYVLANLDSTQASNQLAIFFPFGVAV
THAT TPHENDS TO A CONTROL TO A CONTROL TO A CONTROL THE A CANADA HSDILDEPDFFWDHPETOALYRDVL3CLDIEARINVLIV

681132 681461

yidD family
LYSKMFSMSFKRFLQQIPVRICLLIIYLYQWLISPLLGSCCRFFPSCSHYAEQALKSHGF
LMGCWLSIKRIGKCGPWHPGGIDMVPKTALQEVLEPYQEIDGGDSSHFSE

682494 681391

CT474 hypothetical protein
VLGAKCMAFKRKTRWLWQVLILSVGLNMLFLLLFYSAIFRKDIYKLHLFSGPLIAKSSRK VYLSEDFLNEISQASLDDLISLFKDERYMYGRPIKLWALSVAIASHHIDITPVLSKPLTY TELKGSSVRWLLPNIDLKDFPVILDYLRCHKYPYTSKGLFLLIEKMVQEGWVDEDCLYHF CSTPEFLYLRTLLVGADVQASSVASLARMVIRCGSERFFHFCNEESRTSMISATQRQKVL KSYLDCEESLAALLLLVHDSDVVLHEFCDEDLEKVIRLMPQESPYSQNFFSRLQHSPRRE LACMSTQRVEAPRVQEDQDEEYVVQDGDSLWLIAKRFGIPMDKIIQKNGLNHHRLFFGKV LKLPAKQS

CPn_0594 682517 684958
pheT-phenylalanyl tRNA Synthetase Beta
NTCHYTQVIVKSLVKTSLRLSSMRIPITLLQTYFSEPLSTKEILEACDHIGIEAEIENTT
LYSFASVITAKILHTIPHPNADKLKRVATLTDGEKEHQVVCGAPNCEAGLIVALALPGAKL
FDSEQAYTIKKSKLRGVESQGMCCGADELGLDELQIQERALLELPEATPLGEDLATVLG
NTSLEISLTPNLGHCASFLGLAREICHVTQANLVIPKEFSFENLPTTALDMSNDPDICTP
STYVVTTGISAQPSFIKLGESLQALKQKPINAIVDITNYIMLSLGQPLHAYDASHVALDS
LRVEKLSTPESLTLLNGETVLLPSGVPVVRDDHSLLGLGGGMGAKAPSFQETTTTTVIKA

LRVEKLSTESSITLINGETVILPSGVPVVRDDHSLIGLGGVMGAKAPSFQETTITTVIKA
AYFLPEALRASOKLIPIPSESAYRFTRGIDPONVVPALQAAIHYILEIFPEATISPIYSS
GEICRELKEVALRPKTLQRILGKSFSIEILSQKLQSLGFSTTPQETSLLVKVPSYRHDIN
EEIDLVEEICRTESMNIETONPVSCYTPIYKLKRETAGFLANAGLQEFFTPDLLDPETVA
LTRKEKEEISLQSSKHTTVLRSSLLPGLIKSAATNLINRQAPSVQAFEIGTVYAKHGEQYQ
ETQTLAILLITEDGESRSWLPKPSLSFYSLKGWVERLLYHHHLSIDALTLESSALCEFHPY
QGGVLRIHKQSFATLGQVHPELAKKAQIKHPVFFAELNLDLLCKMLKKTTKLYKPYAIYP
SFFRDLTLTVPEDIPANLLRQKKLHEGSKWLESVTIISIYQDKSLETRNKNVSLRLVFQD
YERTLSNQDIEEEYCRLVALLNELLTDTKGTINS

684943 685926

CT476 hypothetical protein
RDYQFMKQLLFCVCVFAMSCSAYASPRRQDPSVMKETFRNNYGIIVSGQEWVKRGSDGTI
TKVLKNGATLHEVYSGGLLHGEITLTFPHTTALDVVQIYDQGRLVSRKTFFVNGLPSQEE LYNLOGITVLTRWPDNNDSDTITKPYFIETTYQGHVIEGSYTSFNGKYSSSINNGEGVRS
VFSSNNILLSEETFNEGVMVKYTTFYPNRDPESITHYQNGQPHGLRLTYLQGGIPNTIEE
WRYGFQDGTTIVFKXGCKTSEIAYVKGVKEGLELRYNEQEIVAEEVSWRNDFLHGERKIY
AGGIQKHEWYYRGRSVSKAKFERLNAAG

685930 686457

ada-methvltransferase

FAVMADDTLIPKLMKNSLSQACSEGLLIAKYPPLQVIVHFDNNLVVKTHLSVAPVFSCLF LGPAAHKAMQEIVLMCSRYANKEHPPFSSHFAKDLIPSQYLEILNCVAEIPFGEQQTYAE IAKKTDTHPRTVGAACKQNPFLLFFPCHRVVGSHGERNYVLGPVIHEILLKFENSY

CPn_0597 688215 686479

oppC-011gopeptide Permease
MQKHPSFYQRFLSAYYKNLLASLSWKFFISVALIGIYAPLFASSKPLLVTWHGEIFFPLL

RYLFFPGYYTKPVDLFFNVLMVTFPFFILSFKLTRGWLRRWLLGLCIISQCMIFAWAYSG KVQDPALAENLKKMRAEKVRENISKVNSEMVMLLPKDTRTWEMERRYMSTYEQLGILIKA KVODPALAENLKKMRAEKVRENISKYNSEMVMLLPKDTRTWEMERRYMSTYEQLGILIKA
KYRKKQEASVKKYQVAFEEKRQSPMPTLRHLEMKNEGICLKRLQQRVDKMQRPYEMAQQA
WNRATDNYRPFLMALTRIEHELRLADYNNWGQPEDLCIAYANVEKRAEPYKKSLLEIRQV
LEDVAKLRSAISFIQDKRLWIEKESEDLRILINPFFSSFHWEDDAGGSREMNKYVPMVQL
SRVTRKDLLAALVFGIRIALVVAGIGITIALAIGINIGLVSGYFGGTVVDMILSRFTEIWE
TMPVLFILMLVISITOQKSLLLMTVLLGGFSWTGFSRYVRIEVLKQTDMILSRFTEIWE
SHYYIMVHQILPNAIVPVISLVPFAMMAMISCEAGLTFLGLGEESSASWGNLMREGVTGF
PAESAVLWPPAIILTMLLIAIALIGDGVRDALDPRLQDS

688219 689712

CPn_0598 689712 688219

oppB-0ligopeptide Permease

EEGGSVLKYILKRLVLIPLTLFAIVSINFVILNAAPGDVLEEKSRDALGEAGKSDKMRSY
KGPDRYLQFREHYGLTLPIFFNTRPKITHKKIQTALQELANANNTTPSAKNAAKSLVYWG
DCAKFVMPALLFEADDASRDDKYRHIAADLFIRGGVLQGFVGPNLSPEQRAONKEIAESN
AFLVRQLNEEDLDTKVEALKGWFQDHGGTEVFCYSSKOFWKTFFLETRFARTMSRVLRLD
FGTLRNDAHKTVISEVIKRLRGSLVLSILPMIVGFVLCQIFGMIMALKRNRWIDHSLNFI
FLILFSIEVFVAVFWILDNFVINKTIPFTTIFMPYSGLRSPPEVFNELSTLGRIFDLVSH
GELBEGALVSVALAJASHIRGELBIJKSOBELGAAKADGILWHENLIVHKKENBANGILV GFLPFCAVSYGALAAQSRLSRSIFLEVLSQDFICAAKARGLRWFDILYKHVGKNAAVSIV TSLASSLGTLLGGALVVETLFNIDGFGNFFYQAILNRDHNVVLFSVLVGSALSLVGYLLG DICYVLLDPRVQLEGRRI

691823 689682

CPn_0599

oppA-oligopeptide Binding Lipoprotein

KREGSHMYKRCVLPKILKGIVAGSLILLYWSSDLLERDIKSIKGNVRDIQEDIREISRV
VKQQVTSQAIPAARSVMLAPKLYRDEAFALLEGDPSYPNLLSLDPYKQOTLPELLGTNFH
PHGILRTAHVGKRENLSPFNGFDYVVGFYDLCIPSLASPHVGKYEEFSDLAVKIEEHLV
EDGGGGBKEHTYLRENYFMERIDEKALLPHVQLDEVFORH-PHYTAHDIKFFYDAWNNYV
ATMPAVALRICYEDVSVSVSENDLKLVSVFWKAHTVINEEGKEERKVLYSAPSNTLSLQPL
PRFVYOYFANGEKIIEDENIDT/PTNSIWAONFYHMANNYIVGGGAYYFAWNDEKIPF
SRNLDFYDELAALIDKRFVYFYEETTD:LLPDPEKTSKIDICYLPPNGRONTYSFMKSSAYN
KQVAFYSAVRETVSADRAYTYICWNCFELFTD:BUSKCAMBMAIDKERICUCLDCGGYT
FGGPFAGSPBYNKGIFCEWHY PEDETAMELLERS WIEDIKKSIEKKVICH VIVPFRERLC
YYVFSTAHTIADAVYTACKEIGIDCSLIGLEMAEDLKERPHRYBREHEIHERAPYA
FLEGPFAGSBAMEGGANVAGFHREFADFIIDRISTENHARPHEIHERAPYA
FLEGPFAGSLUKKUNVKKHIPVFHPTDLIPFAGDETVNVTMWWLEKKEDPCLDTG

CPn_9600 602156 693827 **C97827** No robust homolog pre-cht in Genebank/EMBL ac of 11 7798

HGYMKIKKSFOYSLCOAKRFONMLPNHFDPCLQPVNLQLKQDRLAYGELIILLSKYQQKT FSSLLKEETCSLNRAKQHLLYKILRDFNTMQHLRSLGLNGWGEIPMSPCL

693092 592736

CT483 hypothetical protein
OFPRIHADDIINSMDEITPNYPLLRODSLWNRYRVSWRADLSVSSRYEIASAIAILGLLV
AFCASAAVSIIFTANPLAQVFIDGCLALGLLPIPLVIGLLIIGIIVLLYGIYLFPQQRE

AFCASAAVSIIFTANPLAQVFELGCLALGELPIPLVIGELIIGIIVLLYGIYLFPQQRE

THE AGENCY OF THE ACT OF TH

694136

hemZ-Ferrochetalase

hem2-Feriochetalase

WKIMRLIVLMQCLVSLFLAKKVTVTTPAYLLANFGGPRHAKDLQEFLISLLTDRDVTGTF
LPRVLHRHLFTFIAKKRVPKVLPQYQSLQNWSPIYFDTETLAKTLSEILRAPVIPFHRYL
PSTHEKTLLALRTLHTRHVIGIPLFPHFTYSVTGSIVRFFMKHVPEIPISWIPQFGSDSK
FVSLITCHIRDFLQKLGILEKECCFLFSVHGLPVRYISQGDPYSKQCYESFSAITTNFKQ
SENFLCFQSKFGPGKMLSPSTAQLCQNIDTDKPNVIVVPFGFISDHLETLYEIERDYLPL
LRSRGYRALRIPAIYSSPLWVSTLVDIVKENSTVVAEELIKSGKKHTGIR

CPn_0604 695981 695196
fliy-Glutamine Binding Protein
CKKRONSEAQLIVYKIKFSWKYNFLICLLAVGLIFFGCSRVKREVLVGRDATWFPKQFGIY
TSDTNAFLNDLVSEINYKENLNINIVNQDWVHLFENLDDKKTQGAFTSVLPTLEMLEHYQ
FSDFILLTGPVLVVAQDSPYQSIEDLKGRLIGVYKFDSSVLVAQNIPDAVISLYQHVPIA
LEALTSNCYDALLAPVIEVTALIETAYKGRLKIISKPLNADGLRLAILKGTMGDLLEGFN AGLVKTRRSGKYDAIKORYRLP

696737 696150

yhhF-Methylase LRKLCSSRGDVRILAGKYKGKSLKTFSNPHIRPTSGLVKEAFFSICREDIEGAAFLDLFA GMGAIGFEALSRGAASVVFVDISIKAIQLIHTNSALLGEQLPVVIFRQDAQSAIQRLIKQ KRSFDLIYIDPPYELCNCYVETLLQKIVSGNILNPEGTLFLENASDEEIACEGLTLRRRR KLGKTYLAEYIVEKDP

CPn 0606 697492 696707

CFME_DOUG 99/492 030/492 CTi参等 hypothetical protein SSYSRRQLRFYTGSLQMHIYGLADLHLALGVPEKTMEVFGDPWIGYHQKICSEMQAVVHP SSTSRRQURFYTGSLOMHTGLADDHLAGAVPERTMEVFGDPWIGYHQXICSEWQAVVHP EDÎYLLPGGISWAMMISEAHKDFAFIGDLORTKYMIGROHDYWSSASTSKILOALPPSLY YLNOGFALLTPHLAVVGVRLWDSPTICVKKENFLTPSTQEQSYTEQDEKIFLRELGRLKR AĒĀĀLPKEVTEVIVMTHYPPISSDGTPGPISEFLEADGRVSLCLFGHIHKVQRPIDGFGN IRGĪHYILVAADYVNFVPQEVM

[1]

GPD_0607 698910 697573

GPD_0607 698910

GPD_0607 6989 PÖKKLFVRDNIIIVPQGTHIPDNYIF

CEn_0608 699690 699016
*UFFdine 5'-Monophosphate Synthase (Ump Synthase)-truncated?
VSFLYFVKNGRRLWRWMNYEDAKLRGQAVAILYQIGAIKFCKHILASGEETPLYVDMRLV
ISSPEVLOTVATLIWRLRPSFNSSLLCGVPYTALITLATSISLKYNIPMVLRRKELQNVDP
SDALKVEGLFTPG GPÜĞİKVSSVFTVPTLIKALIAYGKLSSGDLTLANKISEILEIES

CPH_0609

CPH-0609' 699672 699986
CT490 hypothetical protein
ONTKNSLIRENMLIREFEGISLPKGFPLYLEPPLVLATFQGTQFVGTYSEATNPLYIDNL
NLNYHYTQELLYKAVPCNYKSIYREIPLIIFPEVLIGSTPTQSTE

CPn_0610 701450 700029

rho-Transcription Termination Factor
RIFLEFKGSIMKEERSSEILPRVKETKKHAYVSMQEKSCVGECAVVASESEEAESVTVTK
IAKLORMGIEELMILARQYGVKNIGSLTKSQVVFEIVKAKSERPDELLIGEGVLEVLPDG
FGFLRSPTYNYLPSAEDIYVSPAQIRRFDLKKGDTIIGTIRSPKEKEKYFALLKVDKING
STPDKAKERVLFENLTPLYPMQRIVMEMGKDHLAERVLDLTAPIGKGORGLIVAPPRSGK
TVILOSIAHAIAVNNPDIVLIVLLIDERFEEVTDMIRQVRGEVVASTFDEQPERHIQVAE
MVIEKARRLVEHGNDVVILLDSITRLARAYNTVQPHSGKILTGGVDASALHKPKRFFGAA
RNIEGGGSLTILATALIDTGSRMDEVIFEFFKCTGNMELVLDRRLSDRRTYPAIDLIKSG
TRKEELLYHPSELERVYLFROAIADLTTIDAMHLLIGRIKKTNSNAEFILISLKF TRKEELLYHPSELERVYLFRQAIADLTTIDAMHLLLGRLKKTNSNAEFLLSLKE

702133 701420

CPn_0611 702133 701420
yacE-predicted phosphatase/kinase
RRNRRDAKTSEREDGISYDFIRSYSCEYLMWKKLGPMLKLLKVSITGDLSSGKTEACQVF
QELGAYVVSADEISHSFLIPHTRIGRRVIDLLGSDVVVDGAFDAQAIAAKVFYNSVLLQG
LEAILHPEVCRIIEEQYHQSIQDCNYPLFVAEVPLLYEIHYAKWFDSVILVMANEDIRRE
RFMKKTGRSSEDFDQRCSRFLNVEEKLAQADVVVENNGTKKELHQKIEEYFYALKGAL

704688 702022

CPG_D612 / 70468R 70202
pola-ona-polymetase I
RCHTOLLOTVERPREYAMKKEPVLDASGFIFRAYFALPEMKNHOGOATGAVEGFIRSL
NKLIKEFGERYMIGVFOXPNNKGGROAIYADYKSNPGKKFEDIPPOLALVKEYGSLIGLA
YLEKEGVEADDVIASIAKKAREENYKVYVCTADKDELQLVMDHVVAMMEWADGVVGISE
VIERWITERGNIEPDYLALVGDSSONIFGLFCCOPKKAAALLKOPGSVEGGLERLDAVKGL
SUTMLGEROETLKLGKRLALLDGNIPIPVFIESLTFPOHPVDEEKLIHFYIGGFKTUP
SKGTEAATVDVGIIKOAGSUTNILMLVGCADIAFAVAYTCHHILGIKLEGLALDGGGVF
PIALEEETRYKILFIKADPTIREDUTPYVNIKARBCHALLANGIVIREISYDLALAEHLTN
CKGKLGFGGLLVMHGTTTTAHRFAKGWGNGGLPIGFLPEOPEGYFGETVAYLPIIKDAIL
EETHPPNLNHILGDIEMPLLKVLFGMERAGVPLDVFCLAILGALPETDLAVLTEETYDLS

ORPFNIKSPYQLSDILYNELGLRPIDKAKSTPAEVLEALRSTHPITEKLLAFRTIEKLLS
TYVKALPKOVDSHTQRIHPSFDOTGAVTGRLACRDPNLQNIPTROERGILLRKAFRLSEK
NSYFLSADYSQIELRFLAHLSQDKSLKFAFESGEDIHAFTASOVFHVPLEQVSKEQRMQA
KTVNFGIVYGQQAFGLAKVLKISTGADGELTGAYFSRYPETIAFFVETETIQQAAKDLRVTT
MLGRERIIDSWNEFFGSRAASGRFAVNTRIQGSAAELIKLAMLDISQAIKQQQMKSRMLL
QIHDELLFEVPEEETEEMQRLVREKMESAMTLSVPIVVNILIGKNWAEC

701658 705662 CPn_0613

OND AND A STATE OF THE STATE OF WMMTHAM
KTAPIJAVIEMKOVJASSKYTAKTIONILEGFEKAFLKDRVKGIVIDMOUPGGEVFEIDR
IYSMLRFWKERKGFPIYIYVAGLEASGGYYVSGAATKIYATSSSLIGSIGVRSGPFFNVK
EGLNRYGVESDLLTAGKDKAPMNPYTFWTSHDREERQATLDFLYGGFVDIVTQNRPLLTK
EKLVHTLGARIFSPEKAKQEGYIDVVGATKEQVLQDIVAVCKIEDNYRVIGSGGDWKR VASAAASSPLVTGMIKHDILPLSHDAAYIPPYLAL

707435 705783

CPn_0614 707435 705783
adt-ADP/ATP Translocase
VFIRHKVCKEFMQSSEVKPFSRLRAYLCPIYKSEFSKFVPLFLLAFFVGFNYCLLKNMKD
TLVIVGSDAGAEVIPFLKVWGIVPGAVIVTMVYCWLGSRYPRDTVFYCFMAAFLGFFFLF
AVITYPVCDSLHLNSLADKLQELLPQGLRGFIVMVRYWSYSIYYVMSELWSSVVLSMLFH
GLANQITTITEAGRFYALINTGLNLSSICAGEISYMMGKQTFVAYSFACDSWHSVMLNLT
MLITCSGLIMIWLYRRIHHLTIDTSIPPSRRVLAEEGAATANLKEKKKPKAKARNLFLHL
IQSRYLLGLAIIVLSVNLVIHLFEVVWKDQVSQIYSSHVEFNGYMSRITTLIGVVSVLAA
VLLTQGCIRKWGMTVGALVTPLVMLVSGLLFFGTIFAAKRDISIFGGVLGMTPLALAWT
GGMQNVLSRGTKFTFFDQTKEMAFIPLSPEDKNHGKAAIDGVVSRIGKSGGSLIYQGLLV
IFSSVAASLNVIALVLLIIMVVWIAVVAYIGKEYYSRAADAVATLKQPKEPSSSIVREAQ ESVEQEEMAVL

CPn_0615 708149 707634
pgsA-Glycerol-3-P Phosphatidyltransferase
LAKIMRQFCNLLSLSRLWLALYFCQEKLHIRLLAIVGAMLSDVLDGYLARRYKATSRLGS
LLDPITDKVFVFVCITVLYMEGSLSIAHLFFICARDLFLIIFVCYLSLVKGWKGYDYGSL FWGKIFTVVQFIILLGVTAGGEIPWTGLVPLVALGFLYFLERIMDYKKQFLR

CPn_0616 708704 710137
dnaB-Replicative DNA Helicase
TLITMYESSLIMDKSTGVPLPSPHSKESEMIVLGCMLTGVHYLNLAANQLYEEDFYYLEH
KIIFRVLQDAFKQDKPFIDVHLAGEELKRHNQITVIGGPSYLITLAEFAGTAAYLEEVYDI
IRSKSILRKMISTAKEIEKRALEQPKNVAEALDEAQNSFFKISQSTSVSQYTLVADKLRG LTTTTDREYLVOLQERQELFLQNAQGDMKSFFTGIPTHFIDLDQLIHGFSPSNLMILAAR PAMGKTALALNIAENLCFORRLPIGIFSLEMTVDQLIHRMICSRSEVDSKKISIGDLSGH DFQRIVSVINEMQEHTLIDDQPGLKVSDLRARARRMKESYDIOFLIIDIVLQLLSGSGTL RATESRQTEISEISRMLKTLARELNIPILCLSQLSRKVEDRANHRPMMSDLRESGSIEQD SDLVMFLLRREYYDPNDKPGTAELIIAKNRHGSIGSVPLVFEKELARFRNYSAFECIS

CPn_0617 710481 712316
gldA-FAD-dependent oxidoreductase
LmwthpiaydvivvGaGHAGCEAAYCSAKMGVSVLMLTSNLDTIAKLSCNPAVGGIGKGH
TVREIDALGGIMAEVTDQSGIGFRILNQTKGPAVRAPRAQVDKQLYHIHMKRLLENTPGL
HIMQATVESLLDKEGVISGVTTKEGWMFSGKTVVLSSGTFMRGLIHIGDRNFSGGRLGDP
SSQGLSEDLKKRGFPISRLKTGTPPRLLASSINFSGMEEQPGDLGVGFVHRTEPFQPFLP
OLSCPITHTMEETKALISANLHRSALYGGGIEGVGPRYCPSIEDKIVKFSDKERHHVFLE
PEGLHTQEIYANGLSTSMPFDVQYDMIRSVLGLENAIITRPAYAIEYDYIHGNVIHPTLE
SKLIEGLFLCGGINGTTGYEEAAAQGLIAGINAVNKVFNRPPFIPSRGESYIGVMLDDLT
TOILDEPYMFTGRAEHRLLLRQDNACARLSHYGYELGLISEERYELVKKONQLLEEEKV
RLQKTFRQYGQSVVSLAKALSRPEVSYDMLREAFPNDIRDLGAVLNASLEMEIKYSGYID
RQKILIGSLEKAESLLIPEDLDYKQITALSLEAQEKLAKFTPRTLGSASRISGIASADIQ
VLMIALKKHAHH VLMIALKKHAHH

CPn_0618 712300 713010
lpla-Lipoate-Protein Ligase A
KNMPTTNCIFLDLRGHSILHQLQIEEALLRVANQNFCIINSGAKDSIVLGISRNLNQDVH
ISRAQADHIPIIRRYSGGGTVFIDSNTLMVSWIMNSSEASAQPQELLAWTYGIYSPLLPN
TFSIRENDYVLGHKKIGGNAQYIQRHRWVHHTTFLWDIDLDKLSYYLPIPQQPTYRNQR
SHEEFLTTLRPWFPSRDDFLERIKASGSLLFTWEEFLDNELEEILAQPHRKATTVLN

713462 713013

ndk-Nucleoside-2-P Kinase RRYVYTMEQTLSIIKPDSVSKAHIGEILSIFEQSGLRIAAMKMMHLSQTEAEGFYFVHRE RPFFQELVDFMVSGPVVVLVLEGANAVSRNRELMGATNPAEAASGTIRAKFGESIGVNAV **HGSDTLENAAVEIAYFFSKIEVVNASKPLV**

713519 714145

CPT_U0230
ruvA-H0lliday Junction Helicase
DKMYDYIRGTLTYVHTGAIVIECQGIGYHIAITERWAIECIRALHQDFLVFTHVIFRETE
HLLYGFHSREERECFRILISFSGIGPKLALAILNALPLKVLCSVVRSEDIRALASVSGIG
KKTAEKLMVELKQKLPDLLPLDSRVETSQTHTTSSCLEEGIQALAALGYSKIAAERMIAE
AIKDLPEGSSLTDILPIALKKNFSGVNKD

714144 714707

CFI_0021
ruwf-Crossover Junction Endonuclease
LSRLGSSFKDNKFKVFQESIVSELIIGVDEGTIVAGYAIIAVEGRYQLRPYSYGAIRLSS
DMPLPMRYKTLFEQLSGVLDDTOPONAWLETOFVNKNPQSTMKLAMARGIVLLAAAGRDI
LIFEYAPNVAKKAVVGKGHASKRQVQVMVSKILNVPEVLHPSNEDIADAFALAICHTHVA

715761 714793

CPB_0622 T15761 T14703
CTS03 hypotherical protein
RYSYPLESTEKEHENSENSSSESHYYHSCHRSMEHELGRWKDADIMEWQQICNILSGV
COPMOSKINGLOPETORGHOEBRIHEDYRROLGGALEEEYRRREAKHODLEKLQOENT
WLOMPLAEKLQOIRHONDIIDELPKINLLOGVORTEIJEGRPPLYSHKIKOLEEQLORYVS
OHGAHSITEITEDKOMANYATINPLKKOLIDLOGVOTY [KTYPHSEIAKLREKLORGEGAQ
T)SSYCTIEKUTEVOYTOLAEKKYATALLODIVEDGYCGLPDLHKEKGMAMPSNTKLDHLK
GLIZBYLPDSEVDVVFJESKJEGS

CPG_0623 (17611 71616)
CPG_0623 (17611 71616)

FELWEE: GENET VECKKFGLLPPPAKL (SEV TGCTVVDPVVTSADLNESLQALVRESSDL (NALLGADDA IHFPETEEEPTSASFEESSAMFFPETSSATEEE

CPn_0624 718018 717011
gapA-Glyderoldehyde-3-P Dehyrogenase
AMKVVINGFGIGRLVIRG/LIKRNSSVEVLAINDLVPGDALTYLFKFDSTHGRFPEDVRC
EADHLIVGKRKIGFLSERNVONLPWKDLGVDLVIECTGLFTKKEDAEKHIQAGAKRVLIS
APGKCDIPTEVMCUNHKTENPEKDFYITNASCTTNCLAPIAKVLLDNFSITEGLMTTVHA
ACANDADZIBLIGERI WEGZIRG LOTH I ADJT JAZZAVILCUE ELKGZLTGMAFRVPIEDZ
AVVDLTZREDEGJETYJDI-KZMKZ JOHNELKG LOTTDEQVVOJDE 1936723IFDALAZ
LALNDRFFKLVAWYDNETGYATRIVDLLEYVEKNSK

CPn_0625 718488 718060 rl17-L17 Ribosomal Protein wwqharkkfrvGrtSSHNRCMLANMLKSLIHYERIETTLPKAKELRRHADKMITLAKKNS LAARRIAIGRLMVRYNKLTSKEARQAKGGDTSVYNVDRLVVNKLFDELGNRFVERKGGYT RILKLONRIGDNAQKCIIEFLAS

719670 718495 CPn_0626

CPH_0626 719670 718495

rpoA-RNA Polymerase Alpha
wLpakkkaQSVVLGKEKGMSDNAHNLLYDKFELPEAVKMLPVEGLPIDKHARFIAEPLER
GMGHTLGNALRRALLIGLEAPAIISFANTGVLHEYMAIEGVIEDVTNIILNLKGALLKKY
PMQDSSLGATTQVLKAGISIDASDLAAANGQKEVTLQDLLQEGDFEAVAPDQVIFTVTQP
IQLEVVLRIAFGRGYTPSERIVLEDKGVYEIVLDAAFSPVTLVNYFVEDTRVGQDTDFDR
LVLIVETDGRVTPKEALAFSTQILTKHFSIFENMDEKKIVPEAISIEKENDDILKHL
GGINEIELSVRSTNCLSNANIETIGELVIMPEPRLLQFRNFGKKSLCEIKNKLKEMKLEL
GMDLTQFGVGLDNVKEKMKWYAEKIRAKNTKG

CPn_0627 720059 719640
rs11-Sil Ribosomal Protein
FLIRSRVLVKNQAQAKKSVKRKQLKNIPSGVVHVKATFNNTIVSITDPAGNVISWASAGK
VGYSGSRKSSAFAATVAAQDAAKTAMNSGLKEVEVCLKGTGAGRESAVRALISAGLVVSV IRDETPVPHNGCRPRKRRRV

CPn_0628 720461 720063
rs13-S13 Ribosomal Protein
DAYTILREAGRMPRIIGIDIPAKKKLKISLTYIYGIGSARSDEIIKKLKLDPEARASELT
EEEVGRINSLLQSEYTVEGDLRRRVQSDIKRLIAIHSYRGQRHRLSLPVRGQRTKTNSRT RKGKRKTVAGKKK

721881 720487

sect Translocase
kirtrapymttkofflitelroklfytfalltackvgvfipvpgingelavayfkollg KREFERPYMTILROFFLITELROKLFYTFALLTACRVGVFIPVPGINGELAVAYFKQLLG
SGÖNLFQLADIFSGGAFAQMTVIALGVVPYISASIIVQLFLVFMPALQREMRESSDQKR
RIGRITRLFTVALAVIQSLLFAKFALRMNLTIRGIVLPTLLSSKLFGVPWIFYITTVVVM
TTGTLLLMMIGEQISDKGIGMSISLIIALGILSSFPSVLGSIVNKLNLGSQDSSDLGLIS
IBLEALVFVFVLITTILLIIECVRKIPVQYARRVIGRREVPGGGSYLPLKVNYAGVIPVIF
AŞŞALMFPATIGQFIASESSMMKRIAALLAPGSLVYSICYVLLIIFFTYFWTATQFHPEQ
IASEMKKNNAFIPGIRQGKPTQHYLEYTMNRVTLLGALFLAAIAILPSLLGCLLRVDSNV
SYELGGTAMLIVVGVVLDTMKQVDAFLLMRRYDSVLKTDRTKGRH

CEE_0630 722316 721885 rTTS-L15 Ribosomal Protein MEKESLFDISERKRRKKLLGRGPSSGHGKTSGRGHKGDGSRSGYKRRFGYEGGGVPLYR RVPTRGFSHKRFDKCVEEITTGRLAELFQEGEAITLDALKAKKAIARQAVRVKVILKGDL **EKTFVWQDTAVVLSQGVQNLLGIT**

722812 722312

CEM-19631 722612 72512 72525 Ribosomal Protein EEMSLSKNSHKEDQLEEKVLVVNRCSKVVKGGRKFSFSALILVGDGKGRLGYGFAKANEL TDAIRKGGEAAKKNLMKIEALEDGSIPHEVLVHHDGAQLLLKPAKPGTGIVAGSRIRLIL EMAGIKDIVAKSFGSNNPMNQVKAAFKALTGLSPRKDLLRGAAIND

CPn_0632 723354 722827
rlfe-L18 Ribosomal Protein
KGITSSWLVNLLQVFAPNVLLNLIKVREFVMKMNMSVVKLVKLRKQAKNRSRVMESSLCK
KSIMKRRRALRVRKVLKGSPTKPRLSVVKTNKHIYVQLIDDSIGKTLASVSTLSKLNKSQ
GLTKKNQEVAKVLGTQIAELGKNLQLDRVVFDRGPFKYHGIVSMVADGAREGGLQF

CPn_0633 723760 rl6-L6 Ribosomal Protein 723760 723209

SMSRKAREPILLPQGVEVSIQDDKIIVKGPKGSLTQKSVKEVEITLKDNSIFVHAAPHVV DRPSCMQGLYWALISNMVQGVHLGFEKRLEMIGVGFRASVQGAFLDLSIGVSHPTKIPIP STLQVSVEKNTLISVKGLDKQLVGEFAASIRAKRPPEPYKGKGIRYENEYVRRKAGKAAK

CPn_0634 724215 723787
rs8-S8 Ribosomal Protein
ESSIKRKRIYMOMTSDSIADLLTRIRNALMAEHLYVDVEHSKMREAIVKILKHKGFVAHY
LVKEENRKRAMRVFLQYSDDKKPVIHQLKRVSKPSRRVYVSAAKIPYVFGNMGISVLSTS OGVMEGSLARSKNIGGELLCLVW

724763 724206

rl5-L5 Ribosomal Protein GERKANMSRLKKFYTEEIRKSLFEKFGYANKMQIPVLKKIVLSMGLAEAAKDKNLFQAHL EÉLTMISGQKPLVTKARNSIAGFKLREGQGIGAKVTLRGIRMYDFMDRFCNIVSPRIRDF RGFSNKGDGRGCYSVGLDDQQIFPEINLDRVKRTQGLNITWVTTAQTDDECTTLLELMGL RFKKAQ

CPn_0636 - 725100 724750 r124-L24 Ribosomal Protein PKEKEVMKKONIRVOOKVEILAGODKGKEGKVLOLTEDKVVVEGVNVRIKNIKRSQODK GKRIGIEAPIHISNVRLTIAGEPAKLJVKVTEQGRELWQRRPDGTSQLYRLVRGKKG

THA HIT RITOSOMAL PROFESSI SEMIQUESQUKVADNTGAKKVKCEKVLGGGRRRYATVGDVIVCDVRDVEPNSSIKKGDV FYAVIVETRRHITEKOSSTIKEDTNSCVI IDDKONEKOTE FEGEVARE IRDRGEIKISSL APEVI

Clm_06 38 725771 725490

rg17-g17 Ribosomal Protein NKKEKVKSMASEPPGSRKVKIGVVVSAKMEKTVVVRVERIFSHPQYLKVVRSSKKYYAHT ELKVSEGDKVKIGETRPLSKLKRWRVIEHVGVVS

CPn_0639 725979 725743
r129-L29 Ribosomal Protein
ASGRGINMAAKKOLLTQLRCKSDDDLDAYVHENKKALFALRAENLLQNKVVKVHMFSTHK
KNIARALTVKQEPYGKVHG

file-llb Ribosomal Protein IIMLMPRRTKFRKQQKGQFAGLSKGATFVDFGEYAMQTLERGWVTSRQIEACRVAINRYL KRRGKVWIRIFPDKSVTKKPAETRMGKGKGAPDHWVAVVRPGRILFEVANVSKEDAQDAL RRAAAKLGIKTRFVKRVERV

CPn_0641 727092 726409
rs3-53 Ribosomal Protein
KGRRIMGOKGCPIGFRTGVTKRWRSLWYGNKQEFGKFLIEDVRIRQFLRKKPSCQGAAGF
VVRRNSGKIEVTIQTARPGLVIGKKGAEVDLLKEELRALTGKEVWLEIAEIKRPELNAKL
VADNIARQIERRVSFRRAMKKAMQSVMDAGAVCVKIQVSGRLAGAEIARSEWYKNGRVPL HTLRADIDYATACAETTYGIIGIKVWINLGENSSSTTPNNPAAPSAAA

CPn_0642 727440 727096 rl22-L22 Ribosomal Protein RRHSMFKATARYIRVQPRKARLAAGLMRNLSVQEAEEQLGFSQLKAGRCLKKVLNSAVAN AELHENIKRENLSVTEVRVDAGPVYKRSKSKSRGGRSPILKRTSHLTVIVGEKER

CPn_0643 727725 727450
rs19-S19 Ribosomal Protein
EIRIMGRSLRKGPFVDHHLLKKVRAMNIEEKKTPIKTWSRRSMITPEMIGHTFEVHNGKK
FLTVFVSETMYGHKLGEFSPTRIFKSHPVKKG

CPn_0644 727722 728594

rl2-L2 Ribosomal Protein FIREINSMFKKFKPVTPGTRQLVLPAFDELTTRGELRGTKSKRSLRPNKKLSFFKKSSGG FIRELINSH RAFYIFOLD RELITED THE RESEARCH STREET RESEARCH STREE

728933 728598

728950 729636 CPn 0646

CPH_Ub46 729636 728950

r14-14 Ribosomal Protein

YREDLMVLLSKFDFSGNKIGEVEVADSLFADEGDGLQLIKDYIVAIRANKRQWSACTRNR
SEVSHSTKKPFKQKGTGNARQGCLASPOFRGGGIVFGPKPKFNQHVRINRKERKAAIRLL
LAQKIQTNKLTVVDDTVFVDALTAPKTQSALRFLKDCNVECRSILFIDHLDHVEKNENLR
LSLRNLTAVKGFVYGININGYDLASAHNIVISKKALQELVERLVSETKD

730490 729657 CPn 0647

rl3-L3 Ribosomal Protein
YLEYFSYCKNLPPLITCPFIFLRENFLFFLENSISKILSRFVSLFLQEESKSLLLMDKFM RSHISVMGKKEGMIHIFDKDGSLVACSVIRVEPNVVTQIKTKESDGYFSLQIGAEEMNAP AHTITKRVSKPKLGHLRKAGGRVFRFLKEVRGSEEALNGVSLGDAFGLEVFEDVSSVDVR GISKGKGFQGVMKKFGFRGGPGSHGSGFHRHAGSIGMRSTPGRCFPGSKRPSHMGAENVT VKNLEVIKVDLEKKVLLVKGAIPGARGSIVIVKHSSRT

731636 730605

CPn_0648 731636 730605
CT529 hypothetical protein
FFFKKPCKEVKMATNAIRSAGSAASKMLLPVAKEPAAVSSFAQKGIYCIQQFFTNPGNKL
AKFVGATKSLDKCFKLSKAVSDCVVGSLEEAGCTGDALTSARNAQGMLKTTREVVALANV
LNGAVPSIVNSTORCYQYTRQAFELGSKTKERKTPGEYSKMLLTRGDVLLAASREACTAV
GATTYSAFTGVLRPLMLINKLTAKPFLDKATVGNFGTAVAGIMTINHMAGVAGAVGGIAL
EQKLFKRAKESLYNERCALENQQSQLSGDVILSAERALRKEHVATLKRNVLTLLEKALEL
VVDGVKLIPLPITVACSAAISGALTAASAGIGLYSIWQKTKSGK

CPn_0649 732672 731710 fmt-Methionyl tRNA Formyltransferase LNLKVVYFGTPTFAATVLQDLLHHKIQITAVVTRVDKPQKRSAQLIPSPVKTIALTHGLP LLQPSKASDPQFIEELRAFNADVFIVVAYGAILRQIVLDIPRYGCYNLHAGLLPAYRGAA PIORCIMEGATESCHTVIRMDACMDTGCMANITRVPIGPDMTSGELADALASGGAEVLIK TLQQIESGQLQLVSQDAALATIAPKLSKEEGQVPWDKPAKEAYAHIRGVTPAPGAWTLFS FSEKAPKRLMIRKASLLAEAGRYGAPGTVVVTDRQELAIACSEGAICLHEVQVEGKGSTN SKSFLNGYPAKKLKIVFTLNN

CPn_0650 733513 732665 lpxA-Acyl-Carrier UDP-GlcNAc O-Acyltransferase SRRNMASIHPTAIIEPGAKIGKDVVIEPYVVIKATVTLCDNVVVKSYAYIDGNTTIGKGT SKRWANSTRITTEGALIGUOVIETYVITATILEUVANSTRITUKS TIMPSAMIGNKPODLKYQOEKTYVTIGENCEIREFALITSSTFECTTVSIGNNCLIMPWA HVAHNCTIGNNVVLSNHAQLAGHVQVGGYAILGGWYGVHQFVRIGAHAMVGALSGIRRDV PPYTIGSONPYQLAGIKNKVGLQRGVPFATRIALIKAFKKIYRADGCFFESLEETLEEYG DIPEVKNFIEFCQSPSKRGIERSIDKQALEEESADKEG7LIES

733975 733517

CPI_003t tabZ-Myristoyl-Acyl Carrier Dehydratase MNQPSVIKLRELLDLLPHRYPFLLVDKVLSYDIEARSITAQKNVTINEPFFMGHFPNAPI MPGVLILEALAQAAGVLIGLVLENDRNKRIALFLGIQKAKFRQAVRPGDVLTLQADFSLI SSYCKIKAWAQARVDSQLVTEAELSFALVDKESI

CPn . 9652 734880 733990

CPI_9532 (1488) 733990

Epx. MyElsfoyl Oldnid Deagetylase

KRNTITYODGLNFYMLERTQRTLKREVRYSOVGIHLGKSSTLHLQPAQTNTGIVFQRQG
ARGNYENVPALLPHYYTMRETTLGROGAVIATVERHMAALRSINIIDNLIIOCOGEEIPI
GDOCONDYVELIDQAG ICEGEBKYGIARCHTPPYYYQHQDIFLAAPPODELKIGYTLHYPQ
GOTTOTOYKGLVINEEGIFQGIAACHTACKHELCTLMEKGLIGGGCLDNAVVFKDEGII
GROGLRFADEPVRHKILDLIGDLGLWGRPFVAHVLAVGGGHSSNIAFGKKILEALUL

736499 734868 CPn 9653

cutE-Apolipoprotein N-Acetyltransferase GEPVLRIFCFVIGWCLIAFAQPDLSGFVSILGAACGYGFFWYSLEPLKKPSLPLRTLFVS CFFWIFILEGIHFSWHLSDQYIGKLIYLVWLTLITILSVLFSGFSCLLVAIVRQKRTAFL WSLPGVWVAIEMLRFYGIFSGMSFDYLGWPMTASAYGRQFGGFLGWAQQSFAVIAVNMSF WSLECOWNATIEMENT TO ISOMSE PU LWMMTASAYCROFGOF LWAGOGSFAY LAYMMS YCLLIKKPHAKMLWVLTLLLPYTFGAIHYEYLKHAFQODKRALRVAVVOPAHPPIRPKLK SPIVVWEQLLQLVGPIQQPIDLLIFPEVVVPFGKHRQVYPYESCAHLLSSFAPLPEGKAF LGNGDCATALSOHFQCPVIIGLERWYKKENVLYWYNGSAEVISHKGISVGYDKRILVPGGE VIIMGKFGGLICPPLFFYALGCKRLPGRRSGVZOVRGLPRIGITITYEETFGYRLQSYK AGEARACZICHTUGWZISIG GKVTHFAHGMLPRGI WPF VPACGTGZTAAVGGLBFLK LLYVDTRETKAFJZZCHTJLCHLFAKTLTGGCSLZCHTMLIAFGAVGYZZZGGGGGFLLAK

CPn_0654 737051 736503
vdlD/yciA-acyl-CoA Thioesterase
KKIIDFLSVDRYYRNQEYPJKILSVESTMLKKFPVSFSCIDGHIYKIFPNDLNANNTVFG
GLLMSLLDRLALVVAERHTESVCVTAFVDALRFYAPAYMGENLICKAAVNRTWRTSLEVG
VKVWAENIYKQERRHITSAYFTFVAVNEDNQPIPVHQIVPETPEEKRRYNEADRRRQARL

CPn_0655 737856 737101
dnaQ-DNA Pol III Epsilon Chain
KEIMSLLKDTVFTCLDCEMTGLDVKKDRIIEIAAVRFTFDSVISSIEFLINPERVVSAES
QRVHHISNAMLRDQPKIAEVFPQIKAFFKEDQTIVGHSVGFDLQVLAQEMERIGETFLSK
YTIIDTLALAKEYGDSPNNSLESLAVHFNVPYGNHRAMKDVEININIFKHLCKRFRTLE
QLKQVLAKPIKMKYMPLGKHKGRCFSEIPLAYLQWASKMDFDSDLLFSIRHEIKHRQKGT GFSOVNNPFMEL

CPn 0656 737842 738048

No robust homolog present in Genebank/EMBL as of 11/7/98
THNFLLLPLSLFDILLTVEGFLCLTLYFASVQRMPCEQKRVPGNLYYYYIAAHSSLCLSV CKDTMENKD

CPn_0657 738476 738051
yjeE (ATPase or Kinase)
PMGRYRRVSHSSQETLLLGTELGQVLVPGAVLLLFGDYGAGKTEFVRGIVSGYLGDTIAE
EVASPSFSILHVYGNEFKRLCHYDLYRIDQKNQEYIFQDAEEDDVLCIEWADRLPKPRFC DTINIYITMOTNMEREIIIEKR

CPn_0658 739180 738455

CPIL U058 hypothetical protein
KRWSMDISCAVKOKLLOFLGKOKKPELLATYLFYLEQALSLRPVVFVRDKIIFKTPEDAV
RILEODKKIWRETEIQISSEKPOWNENTKRIYICPFTGKVFADNVYANPODAIYDMLSSC
PONMEKQGGVRIKRFLVSEDPDVIKEYAVPPKEPIIKTVFASAITGKLFHSLPPLLEDFI
SSYLRPMTLEEVQNQTKFQLESSFLSLLQDALVEDKIAAFIESLADDTAFHVYISQWVDT EE

739482 739838

A-Thioredoxin

LESATINOSISTREGKLMVKIISSENFDSFIASGLVLVDFFAEWCGPCRMLTPILENLAA ELPHVTIGKINIDENSKPAETYEVSSIPTLILFKDGNEVARVVGLKDKEFLTNLINKHA

CPS=0660 740327 739860
spot-rrna Methylase
MRYYLHCPDIPQNTGNIGRTCVALGAELILVRPLGFSLADKFVKRAGMDYWDKLQLTVVD
STEEALHDVPEDQIFCLSTKGSASYTEFSLPSSTYVFGSESKGLPKEILKKYYKNCLRI
PMQQDIRSLNLATSVGIVLYEVVRQKTVALQKNPTV

CHALOGO 741139 740327
mip-FKBP-type peptidyl-prolyl c1s-trans isomerase
HSRELKIKDRRRKMNRRWNLVLATVALALSVASCDVRSKDKDKDCGSLVEYKDNKDTNDI
ELSDNQKLSRTFGHLLARQLRKSEDRFFPIAEVAKGLQAELVCKSAPLTETEYEEKMAEV
OKEMPEKKSKENLSLAEKFLKENSKNAGVVEVQPSKLQYKIIKEGAGRAISGKRSALLHY
KGSEINGQVFSSSEGNNEPILLPLQQTIPGFALGMQGMKEGETRVLYIHPDLAYGTAGQL
PENSLLIPEINLIQASADEVAAVPQEGNQGE

742938

CPIE 0662 742938 741172

asps-Asparty1 trna Synthetase
skdermkyrphrcheltsnhigenvolagwhryrnhggvvfidlrdrfgitqivcrede
gpelhgrldavrsewvlsvrgkvcprlagmenphlatghievevasfevlsksonlpfsi
Addhinvneelrleyryldmrrgdiiekllcrhgvmlacrnfmdagffeivtpvlgkst
Addhinvneelrleyryldprrgdiiekllcrhgvgldryfgiatcfrdedlradropef
pegardylvpsriypokfyalpospolfkglunvggldryfgiatcfrdedlradropef
AQIDIEMSFGDTQDLLPIIEQLVATLFATGIEIPLPLAKMTYQEAKDSYGTDKPDLRFD
LKLKDCRDYAKRSSFSIFLDQLAHGGTIKGFCVPGGATMSRKQLGYTEFVKRYGAMGLV
WIKNDEGVASNIAKFMLDEVFHELFAYFDAKDQDILLLIAAPESVANQSLDHLRRLIAK
ERELYSDNQYNFVWITDFPLFSLEDGKIVAEHHPFTAPLEEDIPLLETDPLAVRSSSYDL
VLNGYEIASGSQRIHNDLQSQIFFILKISPESIQEKFGFFIKALSFGTPPHLGIALGLD
RLVMVLTAAESIREVIAFPKTQKASDLMMNAPSEIMSSQLKELSIKVAF

CPn_0663 744220 742901
his3-Histidy1 tRNA Synthetase
KSNHFERRHHVTVTLPKGVFDIFPYLADAKQLWRHTSLWHSVEKAIHTVCMLYGFCEIRT
PIFEKSEVPLHVGEESDVVKKEVYSFLDRKGRSMTLRPEGTAAVVRSFLEHGASHRSDNK
FYYLLPMERYERQOAGRYROHHQFGVEAIGVRHPLRDAEVLALLLWFYSRVGLQHMQIQL
NFLCGSETRFYDKVLRAYLKESMGELSALSQORFSTNVLRILDSKEPEDQEIIRQAPPI
LDYVSDEDLKYPNEILDALRVLEIPYAINPRLVRGLDYYSDLVFEATTTFGEVSYALGGG
GRYDGLISAFGGASLPACGFGVGLERAIQTLLAQKRIEPQFPHKLRLIPMEPDADQFCLE
WOHLRRLGIPTEVDWSHKKVKGALKAASTEQVSFVGLIGERELISQOLVIKNMSLRKEF

CPn_06b4 744775 744557
No robust homolog present in Genebank/EMBL as of 11/7/98
LWCAHAMKKLIALIGIPLVPIKGNTNKEHDAHATVLKAARAKYNLFFVQDVFPVHEVIEP LUPDOLVHYEOWV

744998 746365

THE UNIT OF THE PROPERTY OF TH

ANFCVSLFEIGGLFGMLVAGWLSDKISKGNPGPMNVLFSLGLLFALLGMWFSRSHNOMVV DGTLLFVIGFFLYGFQMMIGLAAAELSHKKAAGTASGFTGWFAYFGATFAGYPLGKVTDV WGWKGFFIALLACASIALLLFLFTWNATEKNTRSKA

746379

GASTANA POL III Alpha
GFFLTWIPLHCHSQYSVLDAMSSIKDFVAKQQEFGIPALALTDHGNLYGAVDFYKECTOK
GIQPLIGGEGYTAPGSPFDYKKEKRSRAAHHLILLCKNEQGYPNLCILTSLAFTBGFYYF
PRITNGLDF/GBFLULL I I I NAF POWALLELWYGLE OF FTEFZLHK
MUBESIAGFMEDWLKUCH LIC MSBESTAGE SEMENCE
EILLINGSGETVRIARONTHIPNEKRKYYRSREYYFKSPAOMAELFKDIPEVISATUMAA
KRCDFTFDFSKKHYPIYVPESLKTLNSYTEEDRYQASAVFLKOLAEEALPKKYSSEVLAH
IAKKFPHRDPIDIVKERMDMEMAIIPREMCDYLLIWMDIHWAKANGIPVGPGRGAG
SVLLFLLGITEIEPIRFDLFFERFINPERLSYPDIDIDICMAGRERVINYAIERHGKDNV SVLLFLLGITEIEPIRFDLFFERFINPERLSYPDIDIDICMAGRERVINYAIERHOKDNV
AQIITFGTMKAKMAVKDVGRTLDMALSKVNHIAKHIPDLNITLISKALETDPDLHQLYIND
AESAQVIDMALCLEGSIRNTGVHAAGVIICGDQLTNHIPICISKDSTMITTOYSMKPVES
VGMLKVDLLGLKTLTSINIAMSAIEKKTGQSLAMATLELDDATTFSLLHQGKTMGIFQME
SKGMQELAKNLRPDLFEEIIAMGALYRFOPMDMIPSFINRKHGKEIIEYDHPLMESILKE
TYGIMVYQEQVMQIAGALASYSLGBGDVLRRAMGKKDFQOMEQEREKFCKRACNNGIDPE
LATVIFDKMEKFAAYGFNKSHAAAYGLITYTTAYLKANYPKEWLAALLTCDSDDIEKIGK
LIREAQSMGIPILPPHINVSSNHFVATDEGIRFAMGAIKGIGRGLIESIVEERDHHGPYE SIRDFIQRSDLKKVSKKSIESLIDAGCFDCFDSNRDLLLASVEPLYFAIAKDKKEAASGV
MTFFILGAMDRKNEVPICLFKDIFTRSKKELLKKEKELLGIYLTEHPMDTVRDHLSKLSV
VLAGEFENLPHGSVVRTVFIIDKVTTKISSKACKFAVLRVSDGIDSYELPIWPDMYEG
QELLEEDRLIYAILVLDKRSDSLRISCRWMKDLSIVVENIIYECDQAFDRIKNQVQKMSF TMSTSGKETKAKGNKPNENGHTQALAPVTLSLDLNELRHSHLCILKKIVQKHPGSRTLVL VFTQDNERVASMSPDDAYFVCEDIEELRQELVTADLPVRVITV

751097 750177

CPH_066/ 75109/ 75017/ GKSFODL

751176 752162

CT547 hypothetical protein
wRFVVVSPRLIMKFLLYVPLLLVLVSTGCDAKPVSFEPFSGKLSTQRFEPQHSAEEYFSQ GQEFLKKGNFRKALLCFGIITHHFPRDILRNQAQYLIGVCYFTQDHPDLADKAFASYLQL PDAEYSEELFQMKYAIAQRFAQGKRKRICRLEGFPKLMNADEDALRIYDEILTAFPSKDL GAQALYSKAALLIVKNDLTEATKTLKKLTLQFPLHILSSEAFVRLSEIYLQQAKKEPHNL QYLHFAKLNEEAMKKQHPNHPLNEVVSANVGAMREHYARGLYATGRFYEKKKAEAANIY YRTAITNYPDTLLVAKCOKRLDRISKHTS

752140 752775

CFIL U0057 (22140 7227) CT548 hypothetical protein IEYLSILPKIEINMRLFSLGTIYLFFSLALSSCCGYSILNSPYHLSSLGKSLLQERIFIA PIKEDPHOQLCSALTYELSKRSFAISGRSSCAGYTLKVELLNGIDKNIGFTYAPNKLGDK THRHFIVSNEGRLSLSAKVQLINNDTQEVLIDQCVARESVDFDFEPDLGTANAHEFALGQ FEMHSEAIKSARRILSIRLAETIAQQVYYDLF

CPn_0670 752738 753196
rsbM-sigma regulatory factor-histidine kinase
PRRLLNRYTMTFFEGETVFPAVLSELHSMLDLIKRAGKQSKCPQEKLLKLELACEELLVN
IISYAYQGENSPOTIAISCISHRGDLEVVIKDHGPSFNPLAVSINIQEDLPLEORKLGGL
GIFLAKSSVDEFLYAREDHCNIVHLKMLNGQHS

753660 753205

CT550 hypothetical protein
RITINQRKYTMSLDFFEEFYHQSILNTGTSFPEGYLNIAEILSYPHCTDANTDFLCSQSD
NDFIIAESKOKLTLFNADFAIWLYDELVQGAVTRGYIAVSQGEGNYEPEMAFEASGQYN
QSSLILEALQLYLKDIKDTENALRSFRFNNDH

753723

CPn_0672 753723 755048
dacf (pbp5) -D-Ala -D-Ala Caroxypeptidase
TIKSPHMKRPFFTYLCIIFYGSCASLSLHAGLSFPEVRGATAAVVHADSGKVFYDKDIDA
VIYPASMTKIATALFILKHYPTVLDTLIKVKQDAIASITPQAKKQSGYRSPPHWLETDGS
TIQLHLREELLGWDLFHALLVCSANDAANVLAMACCGSVEKFMDKLNFFLKEEIGCTHTH
FNNPHGLHHPNHYTTRDLISIMRCALKEPPFRGVISTTSYKIGATNLHGERILSPTNKL
LLPGSTYHYPPALGGKTGTTKTAGKNLIMAAEKNNRLLVTIATGYSGFVSDLYQDVIALC
ETVPNEPLLRKELVPPSDCLQLEIANLGKLSCPLPEGLYYDFYASEDREPLSVSFIAHAD
AFPIEQGDLLGHWVFYDDEGKKISSQPFYAPCRFERTIKPWKLYMKRVFTSYRTYMSITM
ILMYEPTDWLDEVYKNIKHYSKI LLMYFRIRKHRKYKNLKHYSKI

CPn_0673 755242 755463
CT552 hypothetical protein
GKSTEGKAYHCFLKQVSIALNREEVWDNPHHLMFILMQFQQFSGEQDRFGSFLEATIRDR VSFLVLQEKIATLK

CPn 0674 756689

CPn_0674 756689 755577

fmu-RNA Methyltransterase

RGILYVTMYPFROHHAYOLLKOLHTSAISEADRVSYYFKONRSLGSKDROWIONIIFNIL

RHPRILETIILDSGEQVTPEALVAKVNEOVLENLDSYSAIPWPVRYSISDDLAHFLVODY

GEQAEEIAKIWLTEAPITIRVNTDKISVKELQEKLEYPSSPGELPEALHFSKRHPLQST

EAFRRGFFEIODENSORISOJILTDKDIVLDPCAGAGGYSLIFAQKAKHVVINDSRKAI

LQTAKHRLLRAGARNFSLADOLPLGSFOVVIVDAPCGGTGVFRRHPEHKWQFSKKLLLNY

VRVUKDILKQASAYVQPRGRLVYITCGLLKEENEAHVAYMHSLJWKEVHRKTLPLQVGKG DAFFTSHFQKI

CPn_0675 75731 756768
CT696 hypothetical profein
vPLMILDROPSIGNYLRVLELAIRINTPILAYDRKRILLDAWPVNDPLDTNYDTSVGTI
VPLMILDROPSIGNYLRVLELAIRINTPILAYDRKRILLDAWPVNDPLDTNYDTSVGTI
ROVIHELESWGAIGNYLESIGRILLA I FIELPHHELKEGYGWLYRLFFDCYYHIKKAIVDKLCM
PKGLILFEGKRPVDKIVQANK/FDYGYSIN SWDDPTHEVTVSRVGYPTLAGROQRRLAA
DAGLOMIIEALTTLLEGHTAYLFDILELLINGNIGHRKQHJKTLISKGYVTLRELIQLESI
GAEDPOTIIMSIIFOGUSEVLAIRULGRUPHTEBKRQHJKTVCLMQETALASPEDSKIALGEL
AEVLRKVIVEKKLHVSKSDWTTPEEVGHIYTLFKANPALWDKMITMLLMRWLLDYDRDIG
IALEKAAEYYNPHPSFWRQFLPLWQPPP

758051

homologous to CT695 SMGTPTSCMPGERT SMCTPISCNDCDRNTISDPLEESAAEEGDSDLEDRVSESATQVIETIADTGIPEATPSEG TNSDLNSDLVDRVEYEARGSLLTTMLARIRKAVSQIWMHVKTKRHPKEQGVRSLGDIPCD LIKATRIEKETAEPPYTYALETALASCRSFFFHVFLRLFTLLRRCHPEAPLDLCGTDPIS
PEAAVAFALILRSCCKWVATDAVOEGLPLEV IEEAGMYNAFSLEATTTVEEVSKRLSELL
YSDKRIDGLANVRGITKIITSPYLGAGQCVSVVDNLKTYDLGRNYTQVLACASOIDEFAD
KGENFALMKDILYIJRODRSKEI/JDFLMMWSEEHASEVNYDVVLAILEVNLPILEEDYR CHELAYORKLAYVE OF THE PT

CPn_0677 760410 759256

NO robust homolog present in Genebank/EMBL as of 11/7/98
RIAMGINPSGNRSPDDVWVRGAQGDSSTQGTGATNSNLGAHNVTTSTSQPQVASKAKOL
WCTVREFFLGKKSPDSSQGASGPAMQSPSGPTIRPTRPAPPPPTTGGANAKRPATHGKGR
APQPPTAGSSSGSEQPTAMSSEVAKLVSLKDAVHSHAESGKVLKKVSQELQTKWTDWEN
NRGPDVLLHGYRVIARALQCTYTEQSMLIEGTSTGFVPQAVTVAKADAVTQTVRGAIKNL
ENPKPGNDPDGVLMQVVISLGIEGPTLDPGESIQNFLETRVSDFGGDDSDIDYTSDIARL
GSALDRVRENHPNEMPRIWIALARELGAAVHSHATSVRIANAGKNHTRDVVRMANESSRL
LQCMKVLSVGAWANTMTVLIGDLFE

760682 761329

No zobust homolog present in Genebank/EMBL as of 11/7/98
KIIMSVNPSGNSKNDLWITGAHDOHPDVKESGVTSANLGSHRVTASGGROGLLARIKEAV
TGFFSRMSFFRSGAPRAGSQDSBAPSADTVRSDLPGGDARATEGAGRNLIKKGYOFMKVT
IPOVPGGGAQRSSGSTTLKPTRPAPPPPKTGGTNAKRPATHGKGPAPQPPKTGGTNAKRA ATHGKGPAPQPPKGILKQPGQSGTSGKKRVSWSDED

762936 761725

pgk-Phosphoglycerate Kinase
GYMDKLTVQDLSPEEKKVLVRVDFNVPMQDGKILDDIRIRSAMPTINYLLKKHAAVILMS GYMOKLTVODLSPEEKKVLVKVDENVPMOLGKILDDIRIRSAMPTINYLLKARAAVILMS
HLGRPKGOSFOEEYSLOPVVDVLEGYLGHHVPLAPDCVGEVJARQAVQLSPGRVLLLENL
RFHIGEEHPEKDPTFAAELSSYGDFYVNDAFGTSHRKHASVYVVPQAFPGRAAAGLLMEK
ELEFLGRHLLTSPKRPFTAILGGAKISSKIGVIEALLNOVDYLLLAGGMGFTFLQALGKS
LGNSLVEKSALDLARNVLKIAKSRNVTIVLPSDVKAAENLQSKEYSVISIDQGIPPHLQG
FDIGPRITTEEFIRIINQSATVFWNGPVGVYEVPPFDSGSIAIANALGNHPSAVTVVGGGD
AAAVVALAGCSTKVSHVSTGGGASLEFLEQGFLPGTEVLSPSKS

764254 762971

ygod-Phosphate Permease
ysmLpLifrVLLGGFYTSWNIGANDVANAVGPSVGSGVLTLRQAVVIAAIFEFFGALLLG
DRVAGTIESSIVSVTNPMIASGDYMYGMTAALLATGVWLQLASFFGWPVSTTHSIVGAVI
GEGEVLGKGTIIYWNSVGIILISWTLSPFMGGCVAYLIFSFIRRHIFYKNDPVLAMVRVA
PELAALVIMTLGTVMISGGVILKVSSTPWAVSGVLVCGLLSYIITFYYVHTKHCSYISDT PRÄGSLITYRLKERGGNYGRKYLUVERIFAYLQIIVACFMAFAHGSNDVANAIAFVAGVLR QAYPASYTSYTLIRLMAFGGIGLVIGLAIWGWRVIETVGCKITELTPSRGFSVGMGSALT IALBSILGLPISTTHVVVGAVLGIGLARGIRAINLNIIKDIVLSWFITLPAGALLSILFF

CRT 0681 765001 764258
CT691 hypothetical protein
MGFRSHKSFTRSFRQVIIAKKAILMQTLARLFQQSPFAPLQAHLEMVVSCVEYMLPIFTA
LRQGRYEELLEMAKLVSDKEYQADCIKNDMRNHLPAGLFMPISRAGILEIISIQDSIADT
AEBWAILLTIRRLNFYPSMETLFFRFLEKNLEAFELTMTLLHEFNQLLESSFGRKADKA
RLAMGRVAKSEHESDVLQRELMQIFFSDDFIIPEKEFYLWLQVIRRTAGISDSSEKLAHR
INMTLEEK

CFR_0682 764912 765955

dDDD-ABC ATPASE Dipeptide Transport
TSKCLHKNSLFRNNNLPKRSCKRIMASNPILQIEDLSITLAKQRQQYPIVQSLSFTINEG
OTILÆIIGESGSGKSVSAHAILRILECPPFSVSGVNFÇGHNILITASRSIQKKIIGTEISM
IFONPQASLNPVFTIEQOFREIHTHLALTAEVAKEMIYALEETGFHDPRLCLNLYPHQ
LSGGMLQRICIAMALLCSPKLLIADEPTTALDVSVQYQILQLLKTLQKKTGMSLLIITHN
MGVVAETADDVLVLYAGRMVECAPAVQMFHNPSHPYTRDLLASRPSLQPQQLGSFNPIPG
QPPHYTAFPSGCRYHPRCSKILNRCSAEAPEIYPVREGHKVRCWLYDD

CPE_683 765936 766919

dppF-ABC ATPase Dipeptide Transport
GVGENTINFPOPLIQATSLIKHYYKRSFWFOGKTIASRPVDDVSFSLYSRRAVGLIGESG
SGKSTLALAGGLIPLTSGFLIFTNOTPIKLHSKHGRHQLRSQVRLVFQNFQASLNPRKTI
LDSLGHSLLYHKLVPKEKVLATVREYLELVGLSEEYFYRYPHQLSGGQQQRVSIARALLG
VPQLIICDEIVSALDLSIQAQILNMLAELQKKLSLTYLFISHDLAVVRSFCTEVFIMYKG
QIVEKGNTKRIFSDPOHPYTRMLLNAQLFETPDQRQSKPIFQEYHKDSEESCSTGCYFYN
RCPQKQEACKSEIIPNQGDAHHTYRCIH

CPn_9684 768056 767181
spoJ/parB-Chromosome Partitioning Protein
EKSGDIVTEEISKDTIIEVAIDDIRVSPFQPRRVFSNEELQELIASIKAVGLIHPPVVRE
ICTGDRVLYYELIAGERRWRAMQLAGATTIPVILKHVIADGTAAEATLIENIGRVNINNFI
EMAEAFKRLIHVFGLTQDKVAYKVGKRSTVAIKYLALSKTIQESLLQGOITLGHAKV
ILTLEDPILREKLNEIIIQEHLAVREAELIAKQLISEEGSSIELKPTPLDMAESSKQHEE
LQQRLSDLCGYKVQIKTRGSKATVSFHLQNTQDLQKLEAWLSSHGTLSESLS

CPn_0685 76802° 768217
No robust homolog present in Genebank/EMBL as of 11/7/98
FPQSQYLLIFPNRILDLQAFEILDVQGMLTDQRKHIQMLHKHNSIEIFLSNMVVEVKLFF

CPn_0686 768373 758176 No robust homolog present in Genebank/EMBL as of 11/7/98 AKDSMMPGGRLFRYYQELFFFSSYYVCEQRRPRKLYPSLOHLNFPIEKPRFLLKGFKKEL HEYNHY

CPn_0687 768501 763214

CPR_0637 7-85501 - 763214 CT442 hypothetical protein RKIHKMIRHAYRPSTPNCRSPMQKLVHNIWYKFYSFSSAIAICIVLAUFLSIKIVSNTYK HUQAKPHSIELDIRAAEVAVOGGELPSKSAIASIEGAYHUXSESMKPYAGFLALGFYIHN PRIMARYWHIAYMHROQALQIPHFTQKLIKEISEAQAHUXSESMKPXAGFLAHASPE YPTLUFLTVILAWIIAYMHROQALQIPHFTQKLIKEISEAQAHQUTVAVAISKYQLIQTANSSPE YPTLUFLTULRVIELKEILHQDVSODFAALKSSPLFHQFEPMYSDGEWTLSKRFGKKG

169376 770137 CT481 hypothetical protein

SLMLIVLAFRQVFFSHSRSQLDRLKNYLRLLKONFAITLFKERT 'K JHJLMLTFEFASFD FYTNIFPFLEEQYIPAVVGVASRYIPSMAAQDLHPSHRLKEJETLAFJDEIFJNYMFFCC QNELIEMAKSPYIQLASSGFAIRNLMNNPPYLTTEILLJRHHIETITJAKPLAFLFPFGK SDEPSRKLAADHYPYSFLLGNTINRKLKTHNIYRLDIKPMQYVCPSLFQSSRYLKNWIKE KSKQLYLKKQLPKR

771407 770197

CPn_0689 771407 770197

Vfh0-Nif5-related Aminor ransferase
-N*CEPPETZAA AMERICA TABLE TO A NEW AMERICAN TO A TO A NEW AMERICAN TO A TO A NEW AMERICAN TO A TO A NEW AMERICAN GHQCAQPAMERWNVGHVLRVSLGIYNDEDDIDQFILVLQDSLDKIRR

CPn_0690 772704 771436

ABC Transporter Membrane Protein
LSVLRGDKVLVSIETESISISGSPVQKAAEACYTQYSKQPSSKEVLSSFSWIQELSLFPD
RYNLATGASELIKOHWLHNINHSLAFECILINGKYEPSLSQLPEGVIVCGIDEARGSLSSF
MQGFDVNKHPLAFLNAVCSEDRGVVIYIPEEMQTSDPIFVRHISFPTVSDHDVIFSPRIV
VILQQRASAQIQISHDVDLEMVGSSKTIVNGVTELFVGEGADLTVFMVPGYSSEDTLSWS
TIATVEKDAICRMTQNLLESCQGFGWFDNTSVIVOKKGHAESLVLVQSPRKTWVNNLMSH
DAEETVSRQNIKSILYSGHFLFEGTISISSQGDLSDANQKHDTLLLSSEARVSTFPRLEI
ETDEVKASHGATVGPLDPQQIFYMRSRGMTEAEAQEKLIHGFLKQGLVSDTFLGSSFQLN
OTS

773467 772685 CPn 0691

CPn_0691 //346/ //2053
CT691 hypothetical protein
RGLGSMLKIKHLHASCNDVKILDDFNLHIOPGTMHVIMGPNGAGKSTLAKILAGDESVLV
SSGEIALGEQNLLSHLPERSRAGLFVGFQMPPEIPGVNNKMFLRDAYNARRRANQEGDI
SIDEFNTLLSTVLETYEYNATTOLFLDRNVNSGFSGGERKRNEICQMLVLEPEWVLLDEP
DSGLDVDALRLICRVLEKYRELHPTSSLCIVTHNPKLGNLIRPDVVHLLLDGRVALSGDV SLMHELEAKSYQEVTKRVAWR

774945 773461

Transporter ABC Transporter
IQEFCATGLKVMGESVKVFLEEREDYPYGFVTPIESQGLTRGLSEETIEEIAALRNEPQF
IQEFCATGLKVMGESVKVFLEEREDYPYGFVTPIESQGLTRGLSEETIEEIAALRNEPQF
IIDFRLQAYRYMKQLHEPAWARLHYGPIAYDDIVYFSSPKQKKPLGRLEDADPEILDTFK
KLGIPLDEQKRLLNVENVAVDLVPDSVSIGTTFKEALEKAGVIFCSLGEAIQEHPNLVKK
YLGSVVSHRDNFFAALNAAVFSDGSFVYPKGKCEMDISTYFRINNERAGGFERTLIVV
EDGGYASYLEGCTAPAYSSNQLHAAVVELVAHEHAVIRYSTVQNWYAGDKKTGKGGIYNF
VTKRGLCAGYRSKI5WSQVEVGAAITMKYPSCILKGDESVGEFYSVALTSGKMQADTGTK
MLHVGKRTTSTVISKGISSDESKNTFRSLVSLGKKAEHSSNYTQCDSMLIGKASAGYTDP
KIVVENSTSSIEHEATTSKLREDQLLYLRSRGLSPEEAVSLVIHGFCREIIEQLPLEFAQ
EASKLLLIKLENSVG

776292 775240 CPn_0693 775292 775240
TPR Repeats (O-Linked GlcNAc Transferase homolog)
LRSTNHVLGEISMEBAAKHLAKEFLCSGINLFLSGEYEQAEKRLKETLELDSTAALAYCY
LGIIALETGRVSEALNWCSKGLASEPGDSYLRYCYGVALDRGNQYEAAIEQYSAYVALHP
DDVBCWFSLGSVYHRLKRLQBALDCFDKILALDPHNPQSLYNKAVILSEMDBEAESIRLL
EVAVAKNPLYWKAWVKLGFLLSRSKRWDKATEAYERVVQLRPDLSDGHYNLGLCYLTLDK
TRLALKAPQEALFLNAEDADAHFVVGLAHLDLKQMREAYEAFNSALSINLEHERAHYLLG
YLHHMQGETDKATKELLFLQKKDSTFAPLLQKTVVSDPSSMQFERRLDTIS

CPn_0694 779635 776330

pbp2-pbp2-transglycolase/transpeptidase

FSDESEARNIHSMKRPKKFPIYLSIAQKTNRLLSGIVIAFAVIALRLWYLAVVEHEQKLE

EAYKPQIRVLPQYVERATICDRFGKTLAVNQLQYDVSVAYGAIRDLPTRAWRVDEHGHKQ

LIPVRKHYIMCLSELLSQELHLDREAIEDAIHAKASVLGSVPYLVAANVSERTYLKLKML

SKDMPGLHVEAVVRRHYPQESVASDILGYVQPISLQEYKRVTQELSQLRECVRAYEEGED

PKLPEGLASIDQVRALLESVESNAYSLNALVCKMGVEACNDSKLRGKIGKKPILVDRRGN

FIQEMEGAVPEAPGTKLQLTLSASELQAYADALLEFVEKTETFRSAKSLKKREKLPPLPFW

IKGGAIIALDPNNGEILAMASSPRYRNNDFVNAKVAEDSKAVRSSIYRWLENKEHIAEIY

DRKVPLIRERRNPLTGLCYEEILPLTFDCFLDFLFPENSVIKLQLKRNSFVGQAIEVQNL

VTRLLSLFPYEEGTCPCSAIFDAVFPNEEGHTLIQEVISLQEQKWIMECLNQHKADIEEL

KEALDQVFNELPANYDKILYTDILRLIVDPERFSPVLPSEVHRLSLSEFTELQGRVVVLR

SAFSTILEDAFIEVHFKSWRKSEFLQVLAAKRQEALRKGVPTPPVVDYLEEEKTRQYKM

FCQEHLDTFLAYLFSKTPYKEGLEPYYDILDWINELDNGAHRALSWHEHYLFLKERVSH

LSEHLPALFSTFREFNELQRPLLGKYPJSIVRNKRQTEQDLAASFYPVYGKYGYLRPHAYG

QAATLGSIFKLVSAXSVLSQRILWGHNEEPAPNFUVIIDNSFGYRSSKPHVGFFKOGTPI 779635 776330 QAATLGSIFKLVSAYSVLSQRILWGHNEEPANPLVIIDKNSFGYRSSKPHVGFFKDGTPI PTFFRGGSLPGNDFMGRGFIDLVSALEMSSNPYFSLLVGEGLGDPEDLADAASLFGFGEK TGLGLPGEYAGRVPHDLAYNRSGLYATAIGOHTLVVTPLQTAVMLASLVNGGVVYVPKLL LGEWEGEHVSYLSSKKKRTIFMPDAVVEVLKTGMRNVIWGOYGTARAIQSQFPPQLLSRI IGKTSTAESIMRVGLDREYGTMKMKDIWFAAVGFSDQDLSLPTIVVIVYLRLGEFGRDAA PMAVKMIDMWEKIQQRESFLRG

CPn_0695 781382 780201

homologous to CT695

homologous to CT695
SLEVSMKKLLKSALLSAAFAGSVGSLQALPVGNPSDPSLLIDGTIWEGAAGDPCDPCATW
CDAISLRAGFYGDYVFDRILKVDAPKTFSMGAKPTGSAAANYTTAVDRPNPAYNKHLHDA
EWFTNAGFIALNIWDRFDVFCTLGASNGYIRGNSTAFNLVGLFGVKGTTVNANELPNVSL
SNGVVELYTDTSFSWSVGARGALWEGGCATLGAEFGYAGSKPKVEELNVICKVNSGFSVNK
PKGYKGVAFPLFTDAGVATATGTKSATINYHEMQVGASLSYRLNSLVPYIGVGWSRATFD
ADNIRIAQPKLFTAVLNLTAWNPSLLGNATALSTTDSFSDFMQIVSCQINKFKSRKACGV
TVGATLVDADKWSLTAEARLINERAAHVSGQFRF

781703

CPG_0696 781703 782599
CT659 hypothetical plote-in
NSGFYVRMPLITYSNFEIEVOSLEGGSCKLITIKDLMSAGAHFGHOTRRWNPKMKLYIFEE
KNGLY IINLAKTLQQLPHALIHIRKV (QDNKTVLFVCITKKOAKCVIREAAIEAGEFFIAE
RWLZXMLTNMTTIRNSITYTLOK IERDISHANAYITKKEAALLAKHHOKLLENLEGIRYMK
KAPSLLVVVDRSYEKIAVALNKFF7TPVLALVDTNCDFFFIDHVIRCNDDSLKSTRLIIN ATKENTTEVKHKITGTETAGEAKGITEAEDITGVE, ERGÖDDITGDE ENEEEDITVKKELIGEVA

782587 78 1447

Elongation Partor T

STODEGT TADETERALIMOLAGEREN ETRAAKTURKALAGEAGERURKEANTERFEGG ELAKOT LVVKLD ZWELTEANAGEDE AVNIVALGEREAGEIT VID LERERALAGEAGET BADET WOODS JIKLERKETROOG GANTERFE KEVELGEAGERUITERVAATEREN TEVETANGEREN ETTERFORMET KALLES GOTTEN.

TADSLAKDIAMHVVAAQPQFLSKEGVPAEAIAKEKEVIASQIQGKPQEVIEKIVTGKLNT PFQEACLLEQPFIKNADLSIQSLIDDFSKTSGSSVAIEQFILWKIGA

783443 784201

pyrH-UMP Kinase

DYTH-OME KINGSE
EPINKNMAKOTRRVLFKISGEALSKDSSNRIDEMRLSRLVSELRAVRNNDIEIALVIGGGN
ILRGLAEQKELDINRVSADOMOMLATLINGMAVADALKAEDIPGLLTSTLSGEQLADLYT
POKSIEALDOGKILIGTTGAGSPYLTTDTGAALRACELNVOVLIKATMAVDGVVDKDPRL
PHOKUTUDEVLYBERELTIJGTVMAZIZ LIMBJHIGIRVFTFDAG JEZALPOTTOTT VERDVNHVCEDRA

GPn 0699 784179 784721

Trf-Ribosome Releasing Factor
rrf-Ribosome Releasing Factor
TMSVLQDTEKKMAAALDFFHKEVKSFRTCKAHPALVETVVVDVYGTTMRLSDIASISVAD
LRQLVISPYDGNNASAIAKGIIAANLNLQPEVEGSIIRIKVPEPTADYRQEMIKQLRRKC
EEAKINVRNIRREANDKLKKDSALTEDVVKGNEKKIQELTDKFCKQLDELTKQKEAEIAS

CPn_0700 785094 785609 CT676 hypothetical protein LMVHSPTHQCYHCQQPATICYTEIDKDKVIRSYVCATCPCPSHYYNNEHLSLSKGVGVLT LECGNCKTVWHSKQDDEQLLGCHQCYTNFKNQITSKLKSERVVSSSFTMEKGQGSLHIGR APGEASNTNPLLKLIALNEALQDTLEREDYEQAAVIRDQINHLKTKNPDDPS

785584 786672

karg-Arginine Kınase KPKIQMTLPNDLLETLVKRKESPQANKVWPVTTFSLARNLSVSKFLPCLSKEQKLEILQF ITSHFNHIEGFGEFIVLPLKDTPLWOKEFLLEHFLLPYDLVGNPEGEALVVSRSGDFLAA INFODHLVLHGIDFOGNVEKTLDQLVQLDSYLHSKLSFAFSSEFGFLTTNPKNCGTGLKS QCFLHIPALLYSKEFTMLIDEEVELITSSLLLGVTGFPGNIVVLSMRCSLGLTEELLLSS LRITASKLSVAEVAAKKRLSEENSGDLKNLILRSLGLLTHSCQLELKETLDALSWIQLGI DLGLIKVTENHPLWNPLFWQIRRAHLALQKQAEDSRDLQKDTISHLRASVLKELTKGLSP ESF

789700 786929

CPn_0702 789700 786929
yscC/gspD-Yop C/Gen Secretion Protein D
LKKNPVKYYLINIGRKILOGIKKKKKIGILSGLFFLDLVLLGVSSQRPTETSANVKHNL
RDEKLAACPKNSAASLSAKKSHTKKTTPGSIPSKVFSKFDATQDKTFQKTSGSAFPAKPT
TLKELEERKKPRPERRITADVKRSPRFLPTQEVEEPVPAASKEQLDSIQVWEEKQNYARR
AVNAINLSIKKQLEEQTSTVTEKDVQPKYOATPHASKKNVASPSTSMFGIEKAATTVAVP
QDKSEEEKVKERLTKRELTCEDLKDNGYTVNFEDISILELLQFVSKISGTNFVFDSNDLQ
FEVTIVSHDPTSVDDLSTILLQVLKMHDLKVVEQGNNVLIYRNPHLSKLSTVVTDSSLKE
TCEAVVYTRYFRLYSVSPSAAVNI 1QPLLSHDAIVSASEARTHVIISIAGNVDKVSDLL
AÄLDCPGTSVDMTEYEVKYANPAALVSYCODVLGTLAEDDAFQMFIQPGTNKIFVVSSPR
LANKAEQLLKSLDVPEMAHTLDDPASTALALGGTGTTSPKSLRFFMYKLKYQNGEVIANA
LØPIGYNLVYTTAMDEDFINTINSIQWLEVNNSIVIIGNQGNVDRVIGLLNGLDLPPKQV
YIEYLILDTSLEKSWDFGVOWVALGDEQSKVAYASGLLNNTGIATPTKATVPPGTPNPGS
PÜPTPTGQLTGFSDMLNSSSAFGLGIIGNVLSHKGKSFLTLGGLLSALDQDGDTVIVLNP
RIÄQDTQQASFFVGQTVPYGTTNTIIQETGTVTQNIDYEDIGVNLVVTSTVAPNNVVTL
QIĒQTISELHSASGSLTPVTDKTYAATRLQIPPGCFLVMSGHIRDKTTKVVSGVPLLNSI
PĒFERGLFSRTIDQRQKRNIMMFIKPKVISSFEEGTRVTNKEGYRYNWEADEGSMQVAPRH
AĒECGPPSLQAESDFKIIELEAQ APECOGPPSLQAESDFKIIEIEAQ

CPR_0703 791205 789685
pkm5-S/T Protein Kinase
RKIGFMDCRGGIPLPEPQVIGGYHVKKILSKKLRSRVVHGLHPETRHSTVIKVFSPSPSF RKIGFMCCRGGTPLPEPQVIGGYHVKKILSKKLRSRVVHGLHPETRHSTVIKVFSPSPSF TÜRSVVNFLKEAQSLHQITHPNIVKFHRYGKWQDCLYIAMEYIEGISLREYILAQFISLP QAIDIIFDIAQALEHLHSRNILHKDIKPENILITPQXKIKLIDFGLADMDTEIQRAHPSV IĞTPYMSPEQRQGESHSPASDIYALGLLAYELILGHLSLGRVFLSLVPERISKILAKAL QESENNRYSSTREFIQDIHHYRMSGDMQEDLRIKDHTVALYEQLQTQRFWLAPETLRFPD FİSGYLYHQGYPLYPHAYDTLLEGDVFNLWLGYSPISNATIALSVVKSLVQQDLQRPLL DRVGEINECLIRMKIPIDEMGISILCLEISKENKELSWIACGKTVFWIKRQGRVVQDFES FŠPGLGKITSLQIRETKVAMEIGDEAVVCTLELEESVASLKTLSLAELQDRRQKAIFCPI ES HGGIQSRQHGSNSPSTLISLKRIR

792330 791209

flin- Flagellar Motor Switch Domain/YscQ family
RYEMAVAADSSASWLKSRNNFLSSLGKTEEQVAAPEPPKELCQHKIREKFRLEDVQVSIK
FRGSITAVEATKEFGVHLLIQPMVVQPWEVENLLFLTSEEDLQELMVAVFDDASLASYFY
EKDKLLGFHYYFVAEACKLFEELQWVPSLSAKVGGDAIFTATSLQGSFQVVDISLRLDGK NVRCRLLLPEDTFQSCQKFFSGLHDESDLHNIDOTQQISLSVEVGYSQLTQEEMHQVVPG SFIMLDSCLYDPETEESGALLTVQKHQFFGGRFLTPSSGEFKITSYPNLTHEDPPLPENP QASAAPLFGYSBLVVEVARYSLAVSEFIKLNLGSILSLGNHPAYGVDIILDGAKVGRGEI IALGDVLG IRVLEV

CPn_0705 793176 792334
CT671 hypothetical protein
FMELKKTAESLYSAKTINHTVYQNSPEPRDSRDVKVFSLEGKQTRQEKTTSSKGNTRTES
RKFADEERVDDEIAEVGSKEEEQESQEFCLAENAFAGMSLIDIAAAGSAEAVVEVAPIA
VSSIDTQWIENIILSTVESMVISEINGEQLVELVLDASSSVPEAFVGANLTLVQSGQDLS VKFSSFVDATOMAEAADLVTNNPSQLSSLVSALKGHQLTLKEFSVGNLLVQLPKIEEVQT PLHMIASTIRHREEKDQRDQNQKQKQDDKEQDSYKIEEARL

CPn_0706 793689 793180
CT670 hypothetical protein
YAVAKYPLEPVLAIKKDRVDRAEKVVKEKRRLLEIEQEKLREKEAERDKVKNHYMQKIQQ
LRDLLDEGTTSDAVLQIKSYIKVVAVQLSEEEEKVNKQKEVVLAASKELEKAEVNLAKRR KEEEKTRLHKEEWMKEALKEEARAEEKEQDEMGQLLFQLRQKKKRESGGS

CPn 0707 795035 793704

CPn_0707 795035 793704
yoch-7op N (Flugellat-Type ATPase)
WMMOQLTTDFDFTH MOQLADVALLTTVVGR ITEXVVGML (KAVVPNVRVGEVCLVKRNGMEPL
VTEVVGFTO:FARLSPLAELSQUJPSSEVI PTGLPLHIRACHGLLSRVLNGLGEPIDVET
WGFT@FFARLSPLAELSQUJPSSEVI PTGLPLHIRACHGLLSRVLNGLGEPIDVET
WGFT@FFARLSPLAELSQUJPSSEVI PTGLPLHIRACHGLLSRVLNGLGEPIDVET
WGFT@FFARLSQUJESSEVI PTGLPLHIRACHGLLSRVLNGLGEPIDVET
WGFT@FFARLSQUFLN
ALAVYTTA LAEYFRLOGERTVLMMDDVTBFARALREV/LAAVJEPFNRACYTP-VFSTLPRL
LERGGGDDKGT FTAFYTVLNACIDMNEPVADCVK3TLDGHTVL IN \LAQAYHYPA IDVLA
STSRLETATVPEEQRR LIGKAREVLARYKANEML (P. FGCYPPGLORE LOFA IDHIDKLNR
FLKQLTHEKTNYEEAAQQURALFR

CFn_0708

795738

CT668 hypothetical protoin AFKTVKRFFCFMIDEVECFFMLDGDAEAGSITONGGTELAGELKKDIGFFALGSYAAPKD TTLVQGFKPNPMAMGDQNGMLIDPELQEALEGEELQEQINNLKGRLWGFRGTFEDSQTT AQFADEHFQAVGVIIOLINEDLATIAEHTQQDARKEDKEEGSVTRKIIDWYSSGEEVLNR ALLYFSDRDGNRESLANFLK/QYAVQRATQRAELFASIVGTSVSSVKTIMTTQLG

735203 795742

CPn_0709 796203 795743

CThe7 hypotherical profess

COMMAN AMERICAN SERVING STREET MARKET SERVING STREET MARKET SERVING STREET MARKET SERVING STREET MARKET SERVING SERVING STREET MARKET SERVING SERV

796482 796210

CT1666 hypothetical protein
RSRGEKSMATNKSCTAFDFNKMLDGVCTYVKGVQQYLTELETSTQGTVDLGTMFNLQFRM
QILSQYMESVSNILTAVNTEMITMARAVKGS

CPn_0711 796791 796486
CT665 hypothetical protein
TTINNQVLGFINYLYLGRYSMFNMENTAKEEKNSQPLLDLEQDMQDHDRAQELKASVQDK
VHKLHALLREGSDKESFGQQQSLLAGYVALQKVLGRINRKMI

799315 796781

CPn_0712 799315 796781
FHA domain; homology to adenylate cyclase)
MAVRLIVDEGPLSGVIFVLEGG ISMSIGRBSSANDIPIEDPKLGASQAIINKTDGSYYIT
NLDDTIPIVVNGVAIQETTQLKNEDTILLGSNQYSFLSDEFDPQDLVYDFDIPEENFSND
SGDLSDSNEQGKDLEPRGTSETNHSPKPKEKLTKDQSSSDFITSGDQELADAFLASAKAE
KNOPRAKVAKKGLKESSNESLNPKEQNAKDSPKGEERTNKPQNAIMEDNGASPROPOPOP
SAEPSLKNTARDETPLKENKPVEEKANKKATPDSPEKKDOPEGSKKEGSK IEATPLDSQ
KESEDKEAEEAFVQEEEENLTEDNKEDSDSAADANDDTASDHTAEDNKETPKKVENEKSA
VLSPFHVQDLFRFDQTIFPAEIDDIAKKNISVDLTQPSRFLLKVLAGANIGAEFHLDSGK
TYILGTDPTTCDIVFNDLSVSHQHAKITVGNDGGILIEDLDSKWSVIVEGRKIDKTSTLS
SNOVVALGTTLFLLIDHAPADTIVASLSPDDYSLFGRQDQDAEALERGCAQEEEEKQKRA
TLPAGSFILTLFVGGLAILPGIGTASLFHTKEVVPLENIDYQEDLAQVINQFPTVRYTFN
KTNSQLFLIGHVKNSTDKSELLYKVDALSFVKSVDDNVIDDEAWQEMNILLSKRPEFKG
ISMHSPEPGKFIITGYVKTEEQAACLUDYLHIFTYLSLLENKVVETOMLKAILAGHLLQ
GGFANIHVAFVNGEVILTGYVNNDDAEKFRAVVQELSGIPGVRLVKNFAVLLPAEEGIID
LNLRYPNRYRVTGYSRYGEISINVVVNGRILTRGDVIDGMTVTSIQPNAIFLEKEGLKYK
IDYNK

CPn_0713 799817 799332
CT663 hypothetical protein
LDLKEEKAGFRREIVSIPQSTKTTIAALENTSMLEKLIKNFATYMGITSTLELDADGAYV
LPISEVVKYRAQQNADNEIVLSASIGALPPSADTAKLYLQMMIGNLFGRETGGSALGLDS
EGNVMVRRFSGDTTYDDFVRHVESFMNFSETWLSDLGLGKQ

801125 800091 CPn 0714

hemā-Glutamyl trna Reductase
nyrrvimylgvvgisyreaalkereraigylosfeknifiagrfigkggafiplitchra
elyyysespeiagaaliseitsggirpyrhegiscfthlfovtsgidslifgefeiggov
kraylkgskerelpfdlhflfgkalkegkeyrsrigfpdhqvtiesvvgeillsydksiy TNFLFVGYSDINRKVAAYLYQHGYHRITFCSRQQVTAPYRTLSRETLSFRQPYDVIFFGS SESASQFSDLSCESLASIPKRIVFDFNVPRTFLWKETPTGFVYLDIDFISECVQKRLQCT KEGVNKAKLLLTCAAKKQWEIYEKKSSHITQRQISSPRIPSVLSY

801636 803462

CPD_0715 801636 803462

SyrB-DNA Gyrase Subunit B

KPNKISHMAAYTEASILSLASLDHIRLRAGMYIGRLGNGSQKEDGIYTLFKEVVDNGIDE

FIMGHGKSLKISASDKQISIQDQGRGI PLGKLIDCVSKINTGARYTQDVFHFSVGLMGVG

LKAVNALSEIFSVNSVYKKKYHLATFHRGVLQESKQOSTKDPDOTFYSFTDPDSIFPEFT

FNHDFLKDKIRQYTYLHSGLEIRFNDEVFISHNGLKDLFDAEITEPPLYSPLFFQNEDLT

FIFSHLEGNTERYFSFVNQGETLDGGTHLTAFKEAIVKGVNEFFGKTFVSNDIREGIVGC

IAIKIASPIFESGYTNKHLGMTQIRSSLIKDVKEAIVQALRKDKVAPELLLEKIKFNEKTR

KNIOFIKQDLKSKOKKVHYKIPKLRDCKFHYNDRSLYGEASSIFLTEGESASASILASRN

PLTQAVFSLRGKPMNYFSLEETKMYKNDELFYLATALGITQNEIOHLRYNKVILATDADV

DGMHIRNLLITFFLKTLLPLVENNHLFILETFLFKVRNKTTTLYYYSEQEKMQALQQFGK

KDSSLEITRFKGLGEISPKEFAAFIGPEIRLTPVTITSLESISSILOFYMGKNTKERKQF

IMDNLITDF IMDNLITDF

CPn_0716 803466 804902

GyrA-DNA Gyrase Subunit A

FMRDVSELFRTHFMHYASYVILERAIPHILDGLKPVORRLLWTLFLMDDGKMHKVANIAG

FMRDVSELFRTHFMHYASYVILERAIPHILDGLKPVORRLLWTLFLMDDGKMHKVANIAG

RTMALHPHGDAPIVEALVVLANKGYLIDTQGNFGNPLTGDPHAAARYIEARLSPLARETL

FNTDLIAFHDSYDGREKEPDILPAKLPVLLLHGVDGIAVGMTTKIFPHNFAELLKAQIAI

LNDKKFTVFPDFPSGGLMDFSEYODGLGSITLRASIDIINDKTLVVKQICPOSTTETLIR

SIEHAAKRGTIKICTIQDFSTDVPHIEIEIKLPKGSRAKEMLPLLFEHTECGVILLYSKPTVI

YENRPVEGSISEILKLHTTALQGYLEKELLLLOEOLTLDHYHKTLEYTFIKHKLYDSVRE

VLAINKKISADDLHQAVLHALEPWLHELATPVTKQDTSQLASLTIKKILCFNEEACTKEL

LAIEKKQAAIQKDLBRIKEVTVKYLKGLLERHGHLGEPKTQITNFKTAKTSILKQQTLI

CPn_9717 804468 805306 CT6556 hypothetical piotein IRIFFIDTITIVVRMEPRHIYIRKPETPKAPDVEKPGVPEYMTMANTPTFEGPVKTLDQL RRALIEQRGAEEGQKMYDNFIQSILISTFGLVHKDMDPAQKASKRMRSVYKEQ

CPn_0718 R05300 805626
CT657 hyporhetical protein
RA'MSTTYFLALPVDRLMQERFLCSPKRWAPFINSPLYLTLIADHDTTYLAKNLDKFPLP
VECWEKTVLHVSSLLKSIFLASDLSSLRLLACTKFEILTLNDLYGAQNI

806890

LRITMKEFLAYIIKNLVDRPEEVRIKEVQGTHTIIYELJVAKPDIGKIIGKEGRTIKAIR TLLVSVASRNNVRVSLEIMEEK

807671 808489

kdsA-KDO Synthetase

ROSA-KDJ SYNTHELASE KRMYMENNKMILIAGECVIEGEDITLEIAGKLOSILAPYSDRIOWFFKSSYDKANRSSLN SPROPOLTEGLRILAKVKETFGVOILTEVHTPODAYAAAEVCNILO/PAFLCROTDLLVA TAETOA IVNIKKOOFLSPWDMEGGINY/J, STONNKILLITERGCSFG/MINIADMRSIPVL JROSEEVIEGATE //LEFTLTTE //LEFTLTTE //LAFAC DAADMLJEEFAALLITWO/LEFT // JETML /

808477 308974

CRIDUTZZ
200947.
700974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974
701974

808978 809703

CPH_0723 808978 809703
yhbG-ABC Transporter ATPase
ASMPILSVCNLVKKYNKKPVTNDVSFQINPGEIVGLLGPNGAGKTTAFYLTVGLIRPDSG
KIIFKNVDVTKKTMDHRARLGIGYLAQEPTIFKELTVQDNTICILEIIYKARKQQSHLLN
TLVDDLQIGSCLHKKAGTLSGGERRIEIIACVLAINPSVLLLDEPFANVDPLVIQNVKYL
KKLAGRGIGILITDHNAKELLSIADRCYLIIDGKIFFEGSSSQMISNPMVKQHYLGDSF

CPn_0724 810602 809706

No robust homolog present in Genebank/EMBL as of 11/7/98

RTSTRLDYRSGCILSKILPFPELWKMLLGFLCDCPCASWQCAAVANCYDSVFMSRPEHKP
NIPYITKATRRGLRMKTLAYLASLKDARQLAYDFLKDPGSLARLAKALTAPKEALDEGNL
FFYGCSNIEDILEEMRRPHRILLLGFSYCQKPKACPEGRFNDACRYDDSHPTCASCSIGT
MMRLNARRYTTVIIPTFIDIAKHLHTLKKRYPGYQILFAVTACELSLKNFGDYASVMNLK
GVGIRLTGRICNTFKAFKLAERGVKPGVTILEEDGFEVLARILTEYSSAPFPRDFCEIH

810829 810587

CT652.1 hypothetical protein SCGDVGMFFAPLLYESLRRGLMHPTSHMQQQLARLEFINDQLTTELEHVNELLCSLGFPE GLTTIKAIAEEVLSDDEPLLD

813384 810880

CFG_0/26
Tf620 hypothetical protein
ADIDMIYSTSISTFYKKLSLVSSMHSFAQRHRESLEHIANYEKTTAERDILKRLIEVLDQ
RASERYRSAVEKLHKYEVERATVAKSIPVAAIHEKPLSSTHASVQVTASTPAATGSGVGA
YYNAVKOKWAQDLIVELNTYMTIMASVNSKNPANKDVFDKLNTELQALVAAGNNLTEEN YYNAVKOKWAQDLIVELNIVWITIMASVNSKNPANKDVFDKLNTELQALVAAGNNLTEEN
FÜLLYNFPEEIFTAIQRADTFTGGMKTDFTNQLAGKYCNQATLTQTFADGRVEGFKDLLT
AVGGVLTPEQFTIFAEIATELQALADHVCNFDEAGLQRIEDAGEKLAAVINSDLITRNDK
IMPCQHITDLYSDQVAALGSFDTVLDASIYVNQHQGTMFSNLSSFVGSLIGTFAPIDLSS
SQGDISSAALAGALQYTARGLNSRFNELTAEQQKLINECIKSLVTFKCGEHLGAIWAYFTA
STWALNPTATMBHYKAAILEEAKELDNSSFOLASIKSAMTSIVNSSGSFSVTVNSSTL
Q#TIYSEKNGKVEINQILLNYGSTGFLPEITKLAKTNAESTARSYFRFKALAAVESENVQ
NKTEDLQSOLQOFTNMKTELFDGOLLSQASELRALPLPSAVASVLIDRYMPKEVDYLNEI

YRKLYYSNIGSSIGNSIIDAISQYVNGATYFNFASYYGQPAVGAGGANAFPGSQESAQA KLOQERKQAALYLQETRGALTVIEEQRARVLKDDKITNEQRSTILDSLRNYEDNINSISG SIVLQNYLQPLSIAGGSVAGTFEVKEQQEQWQARLQILEEALVSGLVGNMINGGMFPLQ SITQSDQSFADMGQNFQLDLQMHLTSMQQEWTVVATSLQLLNQWYLSLARSITG

_0727 813559 816192

CPn_0727 813559 816192
CT619 hypothetical protein

KYYLFSMSTFSIONRLRTISGESTRIIKLDHKYSGFDPRSVPAINLEELNSGIYALRHLM
NÄLÖSENTNVAALLNPNNTIPPTTSWTDYKHSRPQASSPRAPSSQTPTDIVSAAALALVL
VIDGGLAELVASVTEIDLGALSTISTVROLMASYLGLTTLTAEQEKVVFSSSYVPSEKNL
LEHVKQEKAAEIQAKQEEIKAVLEARGVSTEEIEAILKYPDIYAADFKEFIEEPLHTY
RÄKYGAPIOEMNENAIOLLPTPPAITPDNVNEVNGMNTLSTILQAIDDAIKOAPALGGDQ
EİITILQTLVPLVDKTTFTKAEFDLIYTATQLPNTASLKLYLTDRQIAEYRGKITKVYQN
SLONLSETKRVVENNRSMLETQLSMFQQAQNCFVTWISQANALNIAITNKYISAVLTISM
EMYGGLLCLSYMYERLADDEKAIFDKSVMEYLPIHIVVGGSWVNGWIAKMAAYQELAEYS
LSTÄVTSODOIKAYLOYTRONEFKARTHFFHNIGDMYOFANETVFGNCLTTANGAIQPDL
GGEİREAMTNVGTVEADYVSNAQRILNEFNTAATAHVLQLQLQIAELQKKADDLDPGKAS
FEBRRFAVAAMITSESLGDALISMILNSOLPRQEAFLKPLIEEINFNNLAANALNSLLQ
UTNEFSTTSVYYSLSSYLVQSKTGONLFAGDYYETLLAAAREREYIYRDTARCKQAINLV
NGLLQKINSLPGATSAQKQEMLNATTYYQYSLSVTLNQLTVLESLLAGLKMTLQTTSNNK
YDKSVFKIESFDDMIPTLAALESFLTSGFPNISATGGLGPLFTQVOSDQQTYTSQGQTQQ
LNLQNQMTTIQQEWTUVSTSMQVLMGILSQLAGAIYSN LNLQNQMTTIQQEWTLVSTSMQVLNGILSQLAGAIYSN

CPn_0728 818483 816525
CHLPN 76KDA HOMOLOG (CT622)
VFMVNPIGPGPIDETERTPPADLSAQGLEASAANKSAEAQRIAGAEAKPKESKTDSVERW
SILRSAVNALMSLADKLGIASSNSSSSTSRSADVDSTTATAPTPPPTTFDDYKTQAQTAY
DTIFTSTSLADIQAALVSLQDAVTNIKDTAATDEETAIAAEWETKNADAVKVGAQITELA
KYASDNQAILDSLGKLTSFDLLQAALLDSVANNNKAAELLKEMQDNPVVPGKTPAIAQSL
VDOTDATATQIEKDGNAIRDAYFAGONASGAVENAKSNNSISNIDSAKAAIATAKTQIAE
AQKKPPDSPILQEAEQMVIQAEKDLKNIKPADGSDVPNPGTTVGGSKQQGSSIGSIRVSM
LLDDAENETASILMSGFRQMIHMFNTENPDSQAAQGELAAQARAAKAAGDDSAAAALADA
KKALBAALGKACQQGILNALGQIASAAVVSAGVPPAAASSIGSSVKOLYKTKSKTGSDY
KTQISAGYDAYKSINDAYGRARNDATRDVINNVSTPALTRSVPRARTEARGPEKTDQALA
RVISGNSRTLGDVYSQVSALQSVMQIIQSNPQANNEEIROKLTSAVTKPPQFGYPYVQLS
NDSTQKFIAKLESLFAEGSRTAAEIKALSFETNSLFIQQVLVNIGSLYSGYLQ

819905 818592

CFn_U729 8 19905 818592
CHLPN 76kDa Homolot (CT623)
PAMSOVSTENIOTEKDTMKKQVYQMLASVVLLALTIGGYAELPLGEQKVKSHTYTTLDEVK
DYLGKRGFVETRKODGVLRIAGDVRARWLYFREDIYHSDKDKYNPLPVNRYRGEFYLYI
DYRAERMMLGSKMMWTAIAKGENTAAGVDINRAFLGYRFYKNPETRTDFFMEIGRSGLGD
LPEJBEVQFQSINFOXIAHITWRELDKDYPYQVIVHGGPFVVMMTKKHYAWVVEGILNRLPK
OFFVKGGVVDMNTFVPSETSTTEKAATNAMKYKYLJWWLWINGSQVIWINGGKRPLYLY
GAFIMNPLAKATRTTLMKKRNLAWSIGGTLJGLRVASJAWIATYRYSYVGALGOPEIDVSG TORONGLKEWFAQATAANYDPKEANGETNYKGFGAL/MYGTTD/L.FRAYGAYOKPANDK IJOSDFTFRKEDIJITGAL

CPh_0730 921636 819963 mviN Integral Membrane Protein

DOFKOOGENGLMORKENEVSLARGIENIL JOTEC OF PROTOFICIAN VEYFOADELVAAFW

34

WIFRTVFFLPK LLGGL (LEOAF IPHEFELRAGSLDRAAFFFRFFIRLIK OST IIFTLLIE
AVLWVLQYVEEGTYDMILLTMILLPYGIFLMMYNVNGALLHGENKFFTYGLAFVVVNI I
WIFFYIAARHSDEREP I IGISVALV IGFFFEML ITV PGVWRFLLEAKSPEQERDSVRALL
APLSIG ILTSGIFQILLILGBICLARYVHE IGFLYLMYSLKIYQLFIHLEAGSPETVLLPA
ISRCVQREDHERGLKLMKFVLTLTMSVMI IMTAGLLLLALPGVRVLYSHGLFPQSAVYAI
VRVLRGYGASI IFMALAPLUVLFYAQRQYAVPLFIGIGTALANIVLSLVLRRWYLKVD
GISYATSITAWYQLYFLWYYSSKRLPMYSKLLWESIRRS IKVMGTTMLACMITLGLNILT
OTTYVIFLNPLTPLAWPLGGITAQAIAFLGESCIFLAFLFGFAKLLRVEDLINLASFEYW alle M. TVM.

CPn_U731 821494 821760
No robust nomolog present in Genebank/EMBL as of 11/7/98
VALAISRNIPVIRLQK/PDNILKIERAKETSLSFLLIKPFSPPPLKQDYLFDISPYTSSE
ITIGGSYFKLNKASLQSSTLRRSISIIS

822092 822976

nfo-Endonuclease IV

NFMKVLPPPSIPLIGAHTSTAGGLKNAIYEGRDIGASTVQIFTANQRQWQRRALKEEVIE DFKAALKETDLSYIMSHAGYLINFGAPDPVILEKSRIGIYQEILDCITLGISFVNFHFGA ALKSSKEDCMNKIVSSFSQSAPLFDSSPPLVVLLETTAGQGTLIGSNFEELGYLVQNLKN QIPIGVCVDTCHIFAAGYDITSPQGWEDVLNEFDEYVGLSYLRAFHLDSMFPLGANKDR HAPLGEGY IGKESFKFLMTDERTRKI PKYLETPGGPENWOKE I GELLKFSKNRDS

CPn_0733 823739 823101
rs4-S4 Ribosomal Protein
GLKYMARYCGPKNRVARRFGANIFGRSRNPLLKKPHPPGQHGMQRKKKSDYGLQLEEKQK
KKACYGMIMEKQLVKAFKEVIHKQGNVAQMFLERFECRLDNMVYRMGFAKTIFAAQQLVA
HGHILVNGRRVDRRSFFLRPGMQISLKEKSKRLQSVKDALESKDESSLPSYISLDKTGFK
GELLVSPEQDQIEAQLPLPINISVVCEFLSHRT

823863

YCCA ONTKEHFSSNGNFLQCNYFQDYVRVFIMEKKYYALAYYYITRVDNPHEEIALHKKFLEDL DVSCRIYISEQGINGQFSGYEPHAELYMQWLKERPNFSKIKFKIHHIKENIFPRITVKYR KELAALGCEVDLSKQAKHISPQEWHEKLQENRCLILDVRNNYEWKIGHFDNATLPDIQTF REFPEYAEKLAQECDPETTPVMMYCTGGIRCELYSEVLLEKGFKEVYQLDGGGINAYGOQV GTGKWLGKLFYFDDRLAIPIDESDPDVAPIAECCHCQTPSDAYYNCANTDCNALFLCCDE CIHQHQGCCGEECSQSPRVRKFDSSRGNKPFRRAHLCEISENSESASCCLI

CPn_0735 825680 825003 *Uridine Kinase (Uridine Monophosphokinase) (Pyrimidine

Ribonucleoside Kinase).
GEKFMLMMLMMIGITGGSGAGKTTLTQNIKEIFGEDVSVICQDNYYKDRSHYTPEERAN LIWDHPDAFDNDLLISDIKRLKNNEIVQAPVFDFVLGNRSKTEIETIYPSKVILVEGILV FENQELRDLMDIRIFVDTDADERILRRMVRDVQEQGDSVDCIMSRYLSMVKPMHEKFIEP TRKYADIIVHGNYRQNVVTNILSQKIKNHLENALESDETYYMVNSK

827731 825992

CPH_U/JS 82/592
ygeD-Effiux Protein
RGELLKLARQCLVAFHTVSVKKKSFRALVTTHFLTIINDNLYKFLLAFFLLEGKTLTENA
KILSCVSFFFALPFLLLAPLAGSLADRFQKRNIILATRFIEILCTLIGTYFFFIQSVVGG
YVVLILMACHTTIFGPAKLGILPEMLPSEQLSQANGIMTAATYTGSILGSCLAPLLVDVT YVVLILMACHTTIFGPAKLGILPEMLPSEQLSQANGIMTAATYTGSILGSCLAPLLVDVT
HRLGVNSYVWPTLMCVIVSIISTLISFCIRPSNVKNVKQKITLVSFKDLWKVLKDTRMIH
YLTVSIFLGSFFLLIGAYTQLEIIPFVEFTLKYPKHYGAYLFPIVALGVGTGSYITGKIS
GKDIKIGYVPLAAIGLALVFMGLYAFACSILFVLFFLLALGFLGGVYQVPLHAYVQYASP
EHKRGQILAANNFLDFFGVLVAAGVIRVLGSNLGSPETSFFYIGWFVLAVSIWTLWIWR
EHVYRLLGIILRRQLGYYLKIHQSSSPKCYFVAVQSYREIRRVLAALTKTVRSRYILD
QKLVPGWRAWLLSWCVPTVVSSVRDNDSEAQDAWAVLQANHLKTSLKKFPDVSVVCLGLP
KNVERFTSILQEQGIDLHPIQLVQKEGKKRVIYTLVFPHA

827469 830756

CPn_0737 827469 830756

"reCC-Exodeoxyribonuclease V, Gamma"
KRSAKLPASGASKRKGRAKKKLTQERIFAFSVRVLPSNRKNAKRNLYKLSFIIVRKCVVT
SALNDFFLTETVMNATKHCRASFSNS PRHLLAQLAEDITSTHQKPFTKRWILVANATTGH
WIKNQLVHVLSDHIFMGSTIFTASDSIVKHLFLGSGCSQPNIPDYLTLPLLINNILEEIS
KASKFENGREFLSPPTYETTKKLAAAFKQFHTFSQRPTKNASHYQELFQILESHFSSYEE
MFTTILNNRTQEEDCSLHIFGYAHLPKHLAEFFINLSTYFFVYFYCFSPCREYFGDLLSD
RAIDFFWNQLPDSPINAWEHYVLSDRQALLANLAHKSQSQNFFLDREIDVQEMFLPSK
HDSSLGVIQNSILDLKPTSPQDFSQTKQTICIYRALNIPREVQEVFCKVTELLHRGVSPE
EIFILSSHIESYKVHLNAIFNPHVPIYFTDEVDPRAEDLRNKILLLSSILQTQGDLHYIL
QLLTHPQLQQFICQNKPYPLIKKLSSEXGKISSKDRASGQOMKALGDLILEEYPHQEGG
RVSQVEVWETTVPLIYFIQERINLYLSSSQHSYEDLFQNVFSCLEKIFVLSPEETSFITT
LRNSLFPTFATSSCSLLFFTDFCLDFLLHFHKPSPLYDKFQPYIGSLSSLSLIPKGYVFI
LGANKTISDIFDLLNTTTHEELLFSSTEDEENFHFLQLUSVSKHEHISYISSAAQFN LRNSLFPTFATSSCSLLFTTDFCLDFLLHFHKPSPLYDKPOPYIGSLSSLSLIPKGYVFI
LGANKTTSSDIFDLLNRTTTHEELAFSSTEDEENFHFLQILVSTKHELHISYISSAAQFN
LPSPFLNHIKETLDLEVETLPTOPYLSAFFKNKACLHTSQEYNYSLAHAFYSKKALLPSL
FIPTVKQVNLPQHLSLNEIIKGIFSPLDLFLKTNYNLRISYPEHLKKQQKLPFTKHQIED
FWNECFVDKEHDLIPSISPHAEELFTYYREKTILLRNGLDKDPKHSPYTVTFSSSIFEER
FWHESYLFPPLSLSFQSNPVQIHGTIHGVCNEGLYLCSIDPRDSLKKTTRTLGSLPETSS
EQKQLLERYVALAVLCMSQHLSSDSALIKLTSFNTKENHHPPFSDPEGYLRKVLEVYHLM
SSQPIPLLSPLCWKTLDDEEKFHQAVLSAISEEAKNPSLPIFWQFHNRNIEEILNHVGAS
FRIKIISJERGCFGAV ERLKILSLFRGPCEAV

CPI_0738 833719 833895

"recB-Exodeoxylibonuclease V, Beta"

KFYLFSEVPVKPFNIFDSNSSIQCKFFLEASAGTGKTFTIEQIVLRALIEGSLTHVEHAL
AITFTNASTNELKVRIKDNLAOTLPELKAVLNSQPASLPTYLDINCNVKQIYMQVRNALA
AITFTNASTNELKVRIKDNLAOTLPELKAVLNSQPASLPTYLDINCNVKQIYMQVRNALA
TLEOMSLETHIGGSNSVLEOYFFFTPLIKHKRPALTHSQLYTHHITNYLKQDLKMXNLFQE
QFHLLAVRYNITSKHTSSLVDKLLAGYTOPISSYPSSRVERLEQISLWHQQIYNSLLEIP
KQVFLDQLTNHISGFNKQPFSILDDLHHFVDLLYTSETHISSLFSFFKIAFTFNFKHRLAR
KXFCVAFTVLENNGWYERTLEFKNLLDFIFDDTDKOCWSIFSTNLFISDRSVFALEKL
LSSGEAQPVVQALRFCYQLVLIDFFDDTDKOCWSIFSDNLFISDRSTGSLFICDPKQSIY
EWRSADLPTYLENKSSYPSEDBOKOLGLVNMYRSTFPKLMEAIHQIFGKRSFFLIGDFKQSIY
EWRSADLPTYLTAKSSFSEDBOKOLGLVNMYRSTFPKLMEAIHQIFGKLSFFLIGDFKQSIY
EWRSADLPTYLTAKSSFSEDBOKOLGLVNMYRSTFPKLMEAIHQIFGKUSFFLIGDFKQSIY
EWRSADLFTYLTAKSSFSEDBOKOLGLVNMYRSTFPKLMEAIHQIFGKUSFFLIGDFKCSFCLYDIFGY
EVRTKKERFFTYFFCSHSVLTHFFFFHILTTTALLEAILHFENYGKISFFLIGDFKQSIY
LDTCGSYPYRQLLHLKNRSSFSLWESHLAMFYFYMTPTGSTVLFSSRFGLIFOEMEKLCGY
LDTCGSYPYRQLLHLKNRSSFSLWESHSLANFYFYMTPTGSTVLFSSRSTKELDTHK
DEGGIFFOLFSYSLDRDBGBARTTMINIELLEFFALLTFPRFKIFFSSRSFKELDTHK
DEGGIFFOLFSYSLDRDBGBARTTMINIELLEFFALLTPPRFKIFFSSRSTKFELDTHK
DEGGIFFOLFSYSLDRDBGBARTTMINIELLEFFALLTPPPRFKIFFSSRSTKFELDTHK
DEGGIFFAKLTSSLKRBATTMINIELLEFFALLTPPPRFKIFFSSTKFELDTHK
DEGGIFFAKLTSSLKRBATTMINIELLEFFALTTPPPRFKIFFSSTKFELDTHK
DEGGIFFAKLTSSLKRBATTMINIELLEFFALTTPPPRFKIFFSSTKFELDTHK
DEGGIFFAKLTSSLKRBATTMINIELLEFFALTTPPPRFKIFFSSTKFELDTHK
DEGGIFFAKLTSSLKRBATTMINIELLEFFALTTPPPRFKIFFSSTKFELDTHK
DEGGIFFAKLTSSLKRBATTMINIELLEFFALTTPPPRFKIFFSSTKFELDTHK
DEGGIFFAKLTSSLKRBATTMINIELLEFTALTTPPPRFKIFFSSTKFELDTHK
DEGGIFFAKLTSSLKRBATTMINIELTERGTCSCEPLENGELWCYVIDLEFFEHE
GKYYILDWKTSFLASTNODYSKSHIGHTYIKBFIKTTSLFFACGO

GV (FIRGIUTQONGFFALNSSED (PNFNPKA (OKCQAYH

834892 833861

CT 368 hypothetical protein CKVLFKLMSYSLRNKKTKICVYIIIALGILSFRSIPQEVYDKIRSSFVSLHVKFFPKIKQ CANDERLING FIGHMAN INTLY FITTALD LISENS FROM THE CONTROL OF THE METALL OF THE PRICE

CPn 0740 836054 834864

CPHI_UT40

TyTB-Aromatic AA Aminotransferase
SYMSFFNHIPTFSPDAILGLONVFFADKRPEKVMLVIGVYEHPOKRYGGLSCIRKAQTVI
LEEEONKSYLPISGLQIFLDEMRELVFGAVDPSAIVGFQSLGGTGALHLGARLLSVAKGS
GKVVVPBQTWSNHIRIFSQEGLEVIRYPYYSKEQKQLLFEPLIAFLKEVEKNSVILLHGC CHNPTGVDFTEDMWKELAILMKERELIPFFDTAYQGFAHGIELDRKPIEIFISGRTVLV AASSKNFALYGERVGYFAVHSTFTDELVKIHSFLEEKIRGEYSSPQRWGVEIVSTILSN PYLKEEWSELHFIRESLGKWRTRFVQALRKVAGHTFDFLLSQHGFFAYPGFSDKQVLFL REQHAVYTTAGGRMNLNGITEKNIDHVVQSFIQAYEL

838383 836185

CPn_0741 838383 836185
greA-Transcription Elongation Factor
EYIFFLKTGDIVDYLEKLOVLIEEQCSANFLSLWEEYCFNDVVRGRELVEILEKVKSSSL
ASLFGKIVDTVVPLWEKIPEGKDKDRVLQLILDLQTSNSQMFFDIATEYVNKKYSGEENF
NEALRVVGLRDGRDFQFSLSRFDFLMHMHKGNFVFHQGGWGVGEVMGVSFLQQKVLIEFE
GIMSAKDISFETAFKSLTPLSGDHFLSRRFGDPDGFEAFAKENPIEVVEILLRDLGPKTA
KEIKDELVDLVIPEADMNRWQSAKTKIKKGTRIISPDNPKEPYVLSDAGCSHMQDLERK
EUGLSLNSAEKISLIYHFIRDLHSELKNIEIRKSLVKALQDLDVEEGNKSLILQRELLLSE
YLGIKDASIDKEYITSLSEDDTSRLLENMPIVALQKSFLSLVRKYSSFWQQVFMQILLYT
TSPTMRDFVYKRIIKNDFSSVEVLKKRLLDSAHQPMMFPELFVWFFLKLGNHEDGLFDPED
KEVLRLFLESALNFMYQVASTPHKELGKKLHHYLVGGRYLAVRQMIEGASLPFLKELLLL
STKCPQFSSSDLNVLQSLAEVVQPTLKKHKSNVEEENVLWSTESEFSRMKAKLQSLVGKE
WVDNAKEIEDARSLGDLRENSEYKFALEKRARLQEEIRVLSEEINRARILIKDLVFTDKV GVGCKVTLKGDAGEVVEYTILGPWDADPDSCILSLQSKLAQNMLGKKLNDVVILQGKEYK ISRIQSIWEEHGA

CPn_0742 838442 838888
CT635 hypothetical protein
TKMMVIVMNSKSAQKIIDSIKQILTIYNIDFDPSFGSSLSSDSDADYEYLITKTQEKIQE
LDKRAQEILTGTGMSKEQMEVFANNPDNFSPEEWLALEKVRSSCDEYRKETENLINEITL DEHETKESKRPKOKLSSTKKNKKKNWIPL

NELGFLEVCGEDFALPTLIDPSKTEMLTIVKESLIEYAKESGILTPHQD

841387 840389

CFGL0744 841387 840389
hemB-Porphobilinogen Synthase
EMSSLTLSRRPRRNRKTAAIRDLLAETHLSPKDLIAPFFVKYGNNIKEEIPSLPGVFRWS
LDLLLKEIERLCTYGLRAVMLFPIIPDDLKDAYGSYSSNPKNILCHSIHEIKNAFPHLCL
ISDTALDPYTTHGHDGIFLNGEVLNDESVRIFGNIATLHAEMGADIVAPSDWHDGRIGYI
RSKLDOSGYSKTSIMSYSVKYASCLYSPFRDALSSHVTSGDKKQYQMNPKNVLEALLESS
LDEEGADILMVKPAGLYLDVIYRIRQNTCLPLAAYQVSGEYAMILSAFQQGWLDKETLF
HESLIAIKRAGADMIISYSAPFILELLHQGFEF

CPh_0745 841903 841742

S41/42
NO CODUST homolog present in Genebank/EMBL as of 11/7/98
VDSCPDDWRASSLQGSTTYNVAYDPKHTLAYGFCNQVSVKKFHLKPPKSQEKFL

CPN_B746 · 841939 843567

CT632 hypothetical protein
FSGRCPFSFEVFMLGKEEFTCKOKOCLSHFVTNLTSDVFALKNLPEVVKGALFSKYSRS
VLGRALLLKEFLSNEEDGDVCDEAYDFFTDVQKAADFYORVLDNFGDDSVGELGGAHLA
RENVSILAAKVLEDARIGGSPLEKSTRYYYFDOKVRGGYLYYRDPILMTSAFKDMFLGTC
DFLFDTYSALIPOVRAYFEKLYPKDSKTPASAYATSLRAKVLDCIRGLLPAATLTNLGFF
GNGRFWONLIHKLQGHNLAELRRLGDESLTELMKVIPSFYSRAEPHHHHQAMMQVRRAL
KEQLKGLAEQATFSEEMSSSPSVQLVYGDPDGIYKVAAGFLFPYSNRSLTDLIDYCKKMP
HEDLVQILESSVSARENRRHKSPRGLECVEFGFDILADFGAYRDLQRHRTLTOERQLLST
HHGYNFPVELLDTPMEKSYREAMERANETYNEIVQEFPEEAQYMVPMAYNIRWFFHVNAR
ALQWICELRSQPQGHONYRTIATGLVREVVKFNPMYELFFKFVDYSDIDLGRLNQEMRKE
PTT

843949 844053

CT631 hypothetical protein RTCMGCKGAEVQILSSRSLSGMKILSSSLFYKKFC

844996 844121

CPH_0/48 #44946 #4494 #4494 #4494 #4494 #44946 #44946 #44946 #4494 #4494 #4494 #4494 #4494 #4494 #4494 #4494 #4494 #4494 #4494 #

845638 845006

445004

AND DEPOSITION OF THE PROOF

Cln_0750

846465 845707

tctD/cpxR-HTH Transcriptional Regulatory Protein + Receiver

Doman KITDFILRIHSYNLFCFHMIODKILLFVTEDLSLSSQLKDLASQRSDYQILVSPVFPTSF ESVAIFCEYLLLPBQIFSPJIFPEEDLIVLFDTFQEEAITKVLNQGATGYLLRPITAKVL DAVIRAFLRQHEVLEHSIPDTMTFGDHTFRVLNLVIESPEGSVYLTPSEAGILKKLLINR GHLCLRKNLLAEIKGNTKEIIARNVDVHIASLRKKLGPYGSKIVTIRGVGYLFSDDDSIP LQNHDNTAHPNEE

CEN_A 5 CTh of hypother -416.12 1' 553 110 111 848434

MFRCILFGIFLLTGFSSGG7LYYLFCSHDFSIGPKEKSRSVWIEEEKEFTDSVLHHLPSQ MPRILITOTISCHE THE MONTH OF THE METERS OF THE MONTH OF TH EELLSNKVQQIKTEEAKQCVALLHILDPSISISEKLALSSDTLQNIVSGDDEQHTKLRNY LDLWEAIQSYDIDRQQLVHHLVYGAKDLWKKGGNDEKALNLLQLVLRFTSYDIECESVVF LFIKQAYKQALSSHAIARLIKLEKFISEANIPSIVISEAEKANFLADAEYLFAHEDYDKC YLYSMWLTKVAPSPQSYRLAGLCLMENKRYDEALEFLCMLSPNDSINDYKTQKALAFCQK HOSKDRAAS

CPn_0752 848595 850082

"recD-Exodeoxyribonuclease V, Alpha"

GWALHTEFAPFLEDLVHQQVISPLDIAFASKHISSDFEESFVFLAVSSALWRYGHPFLSL

EENBIRPSLGGISETDLYRGFHNLPKEARDKLFVVVSGRLYLRSLYTIRSKLLDKLSLLC

SATPNYFPPSIDSSILSEEQNFIFNKITGGCFSIVSGGPGTGKTFLAAQLILSLVKQQPK ARIATVSPTGKATSHIRQILMKYNIFDDMVLMGTVHHFLQEYAYRRYNSIDVLLVDEGSM VTFDLLYSLVQTLQGVEKDKKLYYTSSLIILGDTNQLPPIGIGVGNPLQDLIGYFHENTFF LKTSHRAKTGVVDQLTQSVLRGEMISFSPLPSISSAIEVLKNRFVKSLRQSEARLCVLTP MRHGPWGVLNLNTMIHQRLARSDPDLRIPIMVTSRYETWGLFNGDTGLLCLKTQKLHFPQ HEPIDSRALSQYVYYYVMSVHKSQGSEYDEVIVIIPKGSEVFGVSILYTAITRAKYRVSV WGDPETLHKIIKKSNY

851009 850161

No robust homolog present in Genebank/EMBL as of 11/7/98 IMATAHLGRQALLHLRSWTPAIRASGNLFRQQSMSLHNNVLFAGDIVGAIKNSTAISRHA LGSSHYAHAALQKTEGFLGAADGVNTAVAGAMLMGQLLNGSMIFETDEETGELRRCNEAD AEGCMTQKLQRRSALTITGKVARLASKTLGTATFLHEMDVVSLGANANKIGCKVTSCLNL VATGCSLTESSISLYRILSTRPETISDPENRNKPSAEFAARSKAIRNAFIAWLGDVVDLV CDALGTLSLFLPAILGVHAVLIMAILGLISCVINFVKDYAKIG

CPn_0754 851381 851040
rs20-S20 Ribosomal Protein
QFILNLKVLVLSGDIMAPKKPNKKNVIQRRPSAEKRILTAQKRELINHSFKSKVKTIVKK
FEASLKLDDTQATLSNLQSVYSVVDKAVKRGIFKDNKAARIKSKATLKVNARAS

851579 852799

CPH_0755 #S1579 #S2799
CT616 hypothetical protein
YKDLFFMLLVRKWLHTCFKYWIYFLPVVTLLLPLVCYPFLSISQKIYGYFVFTTISSLGW
FFALRRENOLKTAAVQLLQTKIRKLTENNEGLRQIRESLKEHQQESAQLQIQSQKLKNS
LFHLQGLLVKTKGEGQKLETLLLHRTEENRCLKMQVDSLIGEGEKTEEVQTLRELAET
LAYQQALNDEYQATFSEQRNMLDKRQIYJGKLENKVQDLMYEIRNLQLESDIAENIPSQ ESNAVTGNISLQLSSELKKTAFKÆNIEAASSLTASRYLHTDTSVHNYSLECRQLFDSLR EENLGMLFVYARQSQRAVFANALFKTWTGYCAEDFLKFGSDIVISGGKQWMEDLHSSREE CSGRLVIKTKSRGHLPFRYCLMALNKGPLCYHVLGVLYPLHKEVLQS

852889 854676

TPOD-RNA POlymerase Sigma-66
ISYLPLTKLSSKARNPLVLFQVRKLFMNTQNSQATEVSSEESQKKLEELVALAKEQGFI
ISYLPLTKLSSKARNPLVLFQVRKLFMNTQNSQATEVSSEESQKKLEELVALAKEQGFI
TYEEINEILPMSFDTPEQIIQVLIFLTGMDIQVLNQIDVERQKEKKKEAKELEGLARRTE
GTPDDPVRMYLKEMSTVPLLTREEEVEISKRIEKAQVQIERIILRFRYSAKEAISIAHYL GTPDDPVRRYLEMGTVPLLTREEEVEISKRIEKAQVQIERIILRFRYSAKEAISIAHYL ISGKERFDKIISEKEVEDKTHFLKLLPKLITLLKEEDTYLENLLLSLKQPDLSKQEAAKL NDSLEKCRIRTQAYLRCFHCRHNYTEDFGEVVFKAYDSFLHLEQQINDLKVRAERNKFAA AKLAAAKRKLYKREVAAGRTLEEFKKDVRMLQRWMDKSQEAKKEMVESNLRLVISIAKKY TNRGLSFLDLIQEGNMGLMKAVEKFEYRRGYKFSTYATWWIRQAVTRAIADQARTIRIPV HMIETINKVLRGAKKLMMETGKEPTPPELAEELGLIPPDRVREIVKIJOHPISLQAEVGEG SESSFGDFLEDTAVESPAEATGYSMLKDKMKEVLKTLTDRERFVLIHRFGLLDGKPKTLE EVGSAFNVTRERIRQIEAKALRKMRHPIRSKQLRAFLDLLEEEKTGTSKVKSLKSK

854709 855134

folx-Dihydroneopterin Aldolase PCIKNIALVIAIERYOLIISKFRMWLFLGGSVEERHFKOPVLISVTFSYNEVPSACLSDK LSDACCYLEVTSLIEEIANTKPYALIEHLANELFDSLVISFGDKASKIDLEVEKERPPVP NLLNPIKFTISKELCPSPVLSA

855104 856459

856434 856997

fola-Dihydrotolate Reductase LLVRPVIPONERNPLOVENCKNPOVPGTVACDPROVIGLEGKLPWHYPEDLQFFGETIQK FPIVMGRKTWETLEPKYFVDRAVAVFGHEKROVHGEIWVTSLEEFLLLDLSSPTFLIGG GELVOLFLENOIVPDFFSHIYKEYAGDTFFFLGLLCTWTKTVLRPTQFITTCYYENHHS

CPn_07s0 Store; Stored
CTs11 hypotherical profets
RHGPKLCLEIPKPGGPVTMKITT/KTEPTYPVDDLYGILEGGLPKENFFGIVVITSKIVS
LCFGAVVELGKVGPDELIRGEADAZVEZEKYGIYLPPKWGTLIFGSGIDGENVEGYPVLY
PROFELSVATLGEWLKNFYHLGEGGIIIOD:PTTPLRFGTMGLGLZWNGFFPLYNYVGKP
DCFGRALKMTYGNLLDGLSAAAVIX MURYDEGTPIAIIEEAPKITFHGGPPTLQDMGTLA
IAFDEDLYGPLLGGMAWETPARFG

CTelO hypothetical profein

GIMTSWIELDROLEDOHMLKHEFYORWSEGKLEKQQLQAYAKDYYLHIKAFPCYLSALH ARCODLQIRROILENUMDEEAGNPHHIDLWROFALSIGVSEEELANHEFSQAAQDMVATF RRLCDMPQLAVGLGALYTYEIQIPQVCVEKIRGLKEYFGVSARGYAYFTVHQEADIKHAS EEKEMLQTLVGRENPDAVLQGSQEVLDTLWRELSSFINSTEPGSCK

EELLYFSRFSEKQNYSLGNMETKRSIYMALPDRKKALEAAVAYIEKQFGAGSIMSLGRHS
ATHEISTIKTGALSLDLALGIHGVPKGRVIEIFGPESSGKTTLATHIVANAQKMGGVAAY
IDAEHALDPSYASLIGVNIDDLMISQPDCGEDALSIAELLARSGAVDVIVIDSVAALVPK
SELEGDIGDVHVGLQARMSQALRKLTATLSRSQTCAVFINQIREKIGVSFGNPETTTGG
RALKFYSSIRLDIRRIGSIKGSDNSDIGNRIKVKVAKNKLAPPFRIAEFDILFNEGISSA
GCILDLAVEYNIIEKKGSWFNYQEKKLGQGREFVREELKRNRKLFEEIEKRIYDVIAANK TPSVHANETPQEVPAQTVEA

CPn_0763 860520 859972
ygfA-Formyltetrahydrofolate Cycloligase
nFPMTDPKIEKSALRKLFISIRRDLSEERKHEASSAVASFVRSFSKESVVLSFVSFNHEI
DMQEANRILIQKCTLALPKIDQENLYPVLIPSIDDLISVVHPKDPFSKQTPISSDKITHV LVPGLAFDQQGYRLGYGHGFYDRWLAQHPYPSIRTIGIGYCEQKIDRLPQESHDIPLSQI

CPn_0764 861819 860524
CT648 hypothetical protein
GYKSMDIKKLFCLFLCSSLIAMSPIYGKTGDYEKLTLTGINIIDRNGLSETICSKEKLKK
YTKVDFLAPQPYGKVMRMYKNKRGDNVSCLTAYHTNGQIKQYLECLNNRAYGRYREWHVN
GNIKIQAEVIGGIADLHPSAESGWLFDQTTFAYNDEGILEAAIVYEKGLLEGSSVYYHTN
GNIWKECPYHKGYPQCKFLTYTSSGKLKEQNYQCKRHGLSIRYSEDSEEDVLAWEEYH
EGRLLKAEVIDPQTHEIYATIHEGNGIQAIYGKYAVIETRAFYRGEPYGKVTRFDNSGTQ
IVQTYNLLQGARHGEEFFFYDETGRPKLLLNWHEGILMGIVATWYPGGTLESCKELVNNK KSGLLTIYYPEGQIMATEEYDNDLLIKGEYFRPGDRHPYSKIDRGCGTAVFFSSAGTITK

CPn_0765 862415 861801 CT647 hypothetical protein TTIYIKLLGRLMKKWISILILSFLSLLSILPVLAITINHVKISQRWSDLNSQILTLKVIR DHEDQVIKHNARISKDRNNLSIESILNASCKQLRPLSKERERLNKLNSNSLLAQSKEVWER KRÆBEKSNHOLVWNCEOMHNDFAFVRLEOATEMDNEDIESLFSLFNPENPVAPLVFFTCW KMTROTTPLGNEVWLTHAEAISRWI

CPI_0766 863785 862394
CTE46 hypothetical protein
AMNEKLPVYHIGLTKAENNTIKIAILQKTCKGWIVCHCEQIPEGKTWSLPKKYFAAPITF
SE@SDILVKSSSSSLKNRKNILKVALTNLEASLALEWESLIVQPOLGKPTDRGETPLTL
WIAGKNTLKKELSFLSQAQIPPDKLSCRADIFFLAEQSPLKSLPAYLIYGGSEPUTCI
FWEMIAIAVARSFSNHSTKKSCDDIHATLQYIOETFPOTVLPAIHVAQISPNLQKILEQK LSLPLVVCQSMTYGVEDEDWEIYGDTIAAAHHGASRRPLTFPYDATSVSPAAQKHWLLRS SLLTGKYALMATVVVSLGSVLKLKSLSSSASNHFAFACPEEGVLPRSLKAAEKTVKAIGK KNSASNYPLLPTIPTSEOTLKFLLALGKSSPSIKFSYFSYTMTSYPSKDNPSI.PYSAI.VE VKGGGQPEDIPQFLKKISSHPKLQHVSESLEDQRSFKLQFTLSS

CPH_0767 863878 864177
CTF45 hypothetical protein
NIMLSYLLRTAINVYSFLILAYIFASWVPDCQSARWYQLVSKCVDPFLNFFRRFVPRIGF
IDPSPFVGLLCLGILPFVILRVLRFIILNIFHSPWLLQYL

CPn_0768 864144

CPn=0768 864144 865163
yohl/nir3-predicted oxidoreductase
YFSFSMAAPIFIKNILLRSSLYYAPLAGFSDYPYRCMSALYQPGLMFCEMVKVEGILYAP
ERTSKLLDYNENMRPIGAQLCGSNPETSGEAAKILEGLGFDLIDLNCGCPTDKITKDGSG
SGLLKTPELIGRILDKIINSVSIPVTVKIRSGWDMEHINVEDTVRIIRDAGASAVFVHGR
TRAGOYHGPSKQEYISRAKAAAGKEFPVFGNGDIFSPEAAQAMLTTGCDSVLVARGTLGA
PWIGKQIQDYLTTGSYEKIPFIKRKAAFLEHMRLVEDYYQSETKFLSETRKLCGHYLISA
AKYFFLRSSLAKATSYQEVYQLVNDYEEADDSSLETFVKC

CPI_0769 867763 865121

topa-dna Topoisomerase I-Fused to SWI Domain
sigghairlmkkslitvespakiktloklussefypassighivdlpakefgidvdhdf
epoyovlpdkqevinhirklaakcekvylspdpdregealamhlanqlpdspligrysfn
aitknavtealkhprtidmalvnaqoarrlldrivykispilsrklqorsgisagrvqs
valkluvdrekaidafypveywnlrvlmddpkttktfwahlyavogkkwekeipeckten
byllinseekarfyaellekssytitrveakakrrfapppfitstlqqeashffrsasr
misiaqtlyegovldsedstglitymrtdsvrvdpealttvreyiqqtfgkeylpekanv
yttkkmtqdaheairptdinltpdklkiklsddofkvynliwkrfvasqitfaiyotlav
qittdteidlrasgsllkfkgflavyeekqddendpepehplhaqdalikeevsqeq
aftkplprfteaslykeleksgigfpstyatinnkigsryttkenglrptelgkiisq
fletnfprimaligftalmedeleliadnkkfwklllqefwttflpvvitaekeavipril
tniecskchkgklvkiwsknsyfycgsfyecdyrtseellafnkedyaedtpwdspcpt
cogdykkvrhgryttflocgekypecrgtisihkkgeliegeppipcpaicokkifkkrsr
nnkifyscseypecsvignsidavitkysgtekipykkktptikkssakttkaaktpskk
ckakssykkssekktgplflpspdlakmicnepvsrgeatykkidyikehqlqapenkkl
lypdnnlatiigpnpidmfqlskhlsqhltkysndessass

CPn_0770 868322 869131
CT642 hypothetical protein
KPATRNVEKLEFVTSLOSPDDDLITFNKÇOLIAGPEEKVAFLVRSNAMLDAGPETPASF
PESLREOFDIFPEYVEVLYSNEGLDVWEAGCTWILINIEVTIQLRKHHRKASPWLGWYSRD
ES/LAHEAVHAVRMKFHEPVFEEVLAYOTSRWGWRFFFGPLFRSPGESYLLFFTILGLGI
SLWYPAGILIMLVLFMYFLMRLCMAGSYLYRAMKKIPKMLGVPPLWVLLRLTDKEIKMFA
KEPIPVLEHYARKRKLENVRWKQIYGSYFV

870511

CPH_DT71 %0144 EpoH RNA Polymerche Sigma 54 FFYGNCKRIYDSCALDMFQOKQKISIKYIPSIKNQQSLQMLQSPLTELSSYVYQETIDNP FFULSSIKEEEMIR YRPTNITFSYKNOTPSPQESI YTPILPQTEEAFSTAEERFIAHGT ASHISIDE HELMHEDPAQELELPLEKTHKVWTTTQHLSPEGTASPSLQSYWKLIKNSS HQAYSTYPDCYPLMTNCEFAPTMKKFSLSISELRHTLKKALGSTPWCFAAACTYKEMVS TPILPDTYLFYSSSSWKTEVSTRGLBSTKLNKFTFPHFYEHLPKEEQKNLSQQILSAKWLIK

NLRKREGTLLQVMETLLPKQEDFLLGKIPAPYPLGIKDLAEDLJFHEGTIFRAIENKAVA APIGIFPLKHLPPRGIHQDSGHSKENVLQWIRQWIATEQTPLSDGVISDRITAKGIPCAR RTVAKYRAQLKILPANKPKKLFYIRSGNSHFRDRQF

972400 870469 uvrD-DNA Helicase

UVED-UNA HELICASE
KUGLIMICISELNEAQRKAVTAPLNPVLVLAGAGAGKTRVVTYRILHLINQGIAPREILA
VTFINKAARELKEPI'NQCASTNEFDVBMVCTFHSLOVFILRRSINLLINRENNFTIYDQS
DALUURHA LIBERTABLAK LOAF CANBEL HELICANI TOLOMICI CONTYGELIKKL
OFMANJERINLE TTYKLAK FRANFISHLINK NIL PENGETHHAOYTIMOLLIROHP TO ANTI-PER LIGHT THE COME THE STREET, LAKE ALL CHTGOTTHE MOLL ROUP WEAVODDDOSTY JOWEGAN THE LIGHTENDY PHANVLCLEENY RSYGNIL NAANALIKENDA SKLEKELRSVKOFGEK IRLFLGSTDREEADFVAAETLOLHRVORTIKLRDICTFYRTINSOS RTFEDALLRRIPYETIGGLSFYKRKET LODILAFLRIFTSKSDIVAFDRTVNLPKRGTGS TTTFALTDYATAGGLPILKEADODALDTKOVKLSKROOEGLOEYLALFFOTEHAYNTLSLR DFTIESVERTTGYLET LIKEDADTFKORKSNLEELYHKALESOQOPKTHLELFLDDLALKG SDDDLNLTADRVNLMTLHNGKGLEFRVSFLVGLEEQLLPHANSLGGTYENTEERRLCYV GTTRAQDLLYLTAAQVRSLWSTVRMMKPSRFLKETPKDYMIQVR

873195 872485

LPH_U//3
ung-Uracil DNA Glycosylase
FMQNATIOQLPVSWOEQLFLCWREQLKEEWSKPYMQQLLIFLKQEYKEHTVYPEENCVFS
ALRSTPFDQVRVVILGQDPYPGKQQAHGLSFSVPEGQRLPPSLINIFRELKTDLGIENHK
GCLQSWANQGILLINTVLTVRAGEPFSHAGKGWELFTDAIVTKLIQERTHIIFVLWGAAA RKKCELLFNSKHQHAVLSSPHPSPLAAHRGFFGCSHFSKINYLLNKLNKPMINWKLP

873183 873425

CT606.1 hypothetical protein

LEAPMNEGIHSVCFOKTPRLTAKSVVSMEMLLTTQQLPSAEGMPSVANLEADFLRAEALL AEMREIRGCLEOSLRTLVPSE

874040 873414

yggV family
yggV family
ERFMKIVIASSKGYKIRETKTFLKRLGDFDIFSLSDFPDYKLPQEQGDSITANALTKGIH
AANHLGCWIADDTMLRVPALNGLPGPLSANFAGVGAYDKDHRKKLLDLMSSLESLVDRS
AYFECCVVLVSPNQEIFKTYGICEGYISHQEKGSSGFGYDPIFVKYDYKQTFAELSEDVK NOVSHRAKALOKLAPHLOSLFEKHLLTRD

874180 875487

CT605 hypothetical protein FIFVLKNFYDCLLMFFOFLSFTMKKIFYSFVLLSCIFPYVGCAOVFVGLDRIFSEGEYTR CIQGKKIALISHSAAINSRQQDALSVFYSRKHDCTVEILCTLEHGYYGATPTETVGNQPS RYPNLRSVSLYGVKEVPKEVAEHCDVFVYDVQDIGVRSYSFVTVLMQIVKASERYGKQLI VLDRPNPMGGRIVDGPLPNPTTSGSLAIPYCYGMTPGELALFFKKTYAPNANVVVIPMKG WNRSMTFDETGLIWMPTSPQMPDPQSPFFYAATGILGALSVASIGVGYTLPFKVLGAPWM DGEKVADELNRMKLPGVLFLPFFYEPFFGKYKMEMCSGVLLVLQDPKIFYPVETQCTIWG VLKALYPKQVEQTLKSIERIPARRSSICNLFGGDEFLSISHKERYIVWPLRRLCKESRES FHQLRSSCLLSEYAES

875586 877178

groEL_2-heat shock protein-60
TSEDRVVWVFKSOFEGLSALKRGVHALTKAVTPAFGPRGYNVVIKKGKAPIVLTKNGIRI AKEIILQDAFESLGVKLAKEALLKVVEQTGDGSTTALVVIDALFTQGLKGIAAGLDPQEI KAGILLSVEMVYQQLQRQAIELQSPKDVLHVAMVAANHDVTLGTVVATVISQADLKGVFS SKDSGISKTRGLGKRVKSGYLSPYFVTRPETMDVVWEEALVLILSHSLVSLSEELIRYLE LISEQNTHPLVIIAEDFDQNVLRTLILNKLRNGLPVCAVKAPGSRELRQVVLEDLAILTG

ATLIGGESENCEIPVSLDVLGRYKQVMITKETFTFLEGGDAEIQARKQELCLAIARST SESECQELEERLAIFIGSIPQVQITADTDTEQREROPQLESALRATKAAMKGGIVPCGGV AFLRAAHAIEVPANLSCMTFGEFTLLQAVRTPLKVLAQNCGRSSEEVIHTILSHENPRF GYNGMTDTFEDLVDAGICDPLIVTTSSLKCAVSVSCLLLTSSFFISSRTKT

CPn_0778 877400 878092
tsa/ahpC-Thio-specific Antioxidant (TSA) Peroxidase
APVAQSDRVPGYEPGGQRFESSLVRNNKRVEEEVFMTLSLVGKEAPDFVAQAVVNGETCT
VSLKDYLGKYVVLFFYPKDFTYVCPTELHAFQDALGEFHTRGAEVIGCSVDDIATHQQWL
ATKKKQGIEGITYPLLSDEDKVISRSYHVLKPEEELSFRGVFLIDKGGIIRHLVVNDLP
LGRSIEEELRTLDALIFFETNGLVCPANWHEGERAMAPNEEGLQNYFGTID

878502 878095

CT602 hypothetical protein
RFDLIFOMKFTVALFGEAEKGSYDTAYFCRSLVDLHNYLGDVSSPGITLAIKTLLSDYNV
VYFRVREEGYCVDSYFFGLHFLNTQTTLKNIIAIGLPGVGNQHIIEASRSLCQKHNSLLL FFDHDLYDLLTFNOPF

CPn_0780 879241 878591
papQ/amiB-N-Acetylmuramoyl-L-Ala Amidase
HGNRIAVQSLRRMHAKLSFFILLSLLFSGIDCSRLHAAGRSPSLQGVLAEIEDISAKLAS
HEVEIVMLSERLDEQDSKCQKWTAAKPETLAQKIRELESDQKALAKTLAVLTTSVKDLQT
NLQSKLQEIQKDHRALAQDLRLVRRSLLALVDSSSFGAYADFSDPVPENIYIVREGDSLS
KIAKKYKLSVTELKKINKLDSDAIYAGQRLCLQPNKQ

879851 879198

pal-Peptidoglycan-Associated Lipoprotein QNCYRSRRKTVPLLGCFPSATDKENTMNIHSLWKLCTLLALLALPACSLSPNYGWEDSCN TCHHTRRKKPSSFGFVPLYTEEDFNPNFTFGEYDSKEEK/YKSSQVAAFRNTFATDSYT IKGEENLAILTNLVHYMKKNPKATLYIEGHTDEPGAASYNLALGARRANAIKEHLRKQGI SADRLSTISYGKEHPLNSGHNELAWQQNRRTEFKIHAR

CPn 0782 881077 879773

POWGAFFOORTKRIPL

AKT IRP::VATOPQKQAKCSPPQENVQKALQKPIPKVIKTEPPKPSPAPTVAKKTTATEKP PPSTTKKNTQLSKTQLQTLSEVAQALSLHVDKIEKSETSLKNISWPSTAQLTMHSELKAT QEDELCELFRTHIALPSKGYVRIKLVLSPNGEIQECSFLSEVSAADKQLLTQRIQALPFQ KFLEKYKVSKNISFHIKLVONES

CPn_0784 882359 891892
exbD-Biopolymer Transport Protein
DRADSIETTERPYSOIPPYKIMKYPTTBEIEEEPLVNLTPLIDIVFVILMAFIVAVPLIK
ENNIALAB TORK, FVL. IMB. JAZI- Z. ABB LITZBEB. TOOLUTZBETLIBEA FV.
TPLILQLOBTSBEIY/TZKNAIRAZ BEBLHVALCE

883039 882296

exbB/tolQ-polysaccharide transporter
DHLYFETLSVNKDFYSMVHFSHNPIIQAYTEADFFGKSIFFCLLILSVCTWTVLHQKLAI OKNPLKACKSLKDFLIKNRHAPLSLDIHPELSPFADLYFTIKRGTLELLDKNROSAPDRG PILSSEDIOSLETLLGAIMPKYKALLHKNSFIPATTISLAPFLGLLGTVWGILVAFTHIS SGSSGNSAIMEGLATALGTTIIGLFVAIPSLIAFNYLKAHSSELISEIEQTAYLLLNSIE VKYRNTNL

CPn_0786 883137 885293
dsbD/xpra-Thio:disulfide Interchange Protein
NHGVILNKFKTYLQTALIAPFFSFPALSGSFSSIQAEEITQQVMHPGAELLSEGSYIPGL
QTFRLGIKITASKGSHIYWKNPGEIGSPLKISWQIPKGFVVEEEHWPTPKVFEEEGTTFF
GYEDSALIVADVRAPEGYTPGQEVELRAQVEWLACGDSCLPGNVOLKLTLPYSEKEPSLY
PDTHAEFTKTLHAQPRVLENDHSVQVAQCKGNEIILNISKKINATKAWFVSEKADKLFAY
AETSYSGGTGTAWRLKVKNLSGVQKNEKLHGILLLADHTGRPVESLTIHSEVLQQTGSAV
AGLSQYITILIMAFLGGYLLNIMPCVLPLVTLKVYGLIKSAGEHRSSVIANGLWFTLGVV
GCFWGLAGVAFILKVLGHNIGWGFGLQEPMFVATLIIVFFLFALSSLGLFEMGTMFANLG
GKLQSSEMKSSNNKAVGAFFNGILATLVTTPCTGPFLGSVLGLVMSLSFLQQLLIFTAIG
GKAGSSEMKSSNNKAVGAFFNGILATLVTTPCTGPFLGSVLGLVMSLSFLQQLLIFTAIG
GMASPYLVFSVPFKMLSVLEKPGGWMSTFKQLTGFMLLVTVTWLVWIFGSETSTTSVVV
LLGGLWLAGLGAWILGRWGTPVSPKKQRVCASLLFFAFLGGAISVSGLASHYFAEPQGTV
SVNEDSLWQPFSLEKLAQLRAQGRPVFVNFTAKWCLTCQMNKFVLYGDAVQKMFETHGIV
TLEADWTRKDPGITEELARLGRASVPSYYYYPGDNSAPVVLPEKTIQNLLEDVVSNFVR TLEADWTRKDPGITEELARLGRASVPSYVYYPGDNSAPVVLPEKITQNLLEDVVSRFVR

CPn_0787 885604 886401
yabD/ycfH-PHP superfamily (urease/pyrimidinase) hydrolase
TRRQPVDLADAHVHLSDDAFEEDINSVLQRAQDSGVSLVVNVTTTEKELNRSFAYAERFP
KIRFCHVGGTPPQDVDQDIEEDYRNFHAAAHSKLAAIGEVGLDVCFATEEJIARQKEVL
QRYLALSLECLELLVVHCRGAFNDFFRMLDQYYHNDPRSREGMLHCFTGTLEEAQELISR
GWFISISGIVTFKNAQDLRDLVVELPLEHLLIETDAPFLAPVPYRGKKNEPAHVLHTINA
VANNKGMFPQELAALAYKNVLRFLHG

CPm 0788 887432 886521

CHI_30766
sdiTc-Succinate Dehydrogenase
SLIKSLRMSRHEICPEVSHKKGKYYSTFIFRCIHSLAGIAFTFFLCEHLFTMLASSYFS
CGKGFVAMVMGPHKIPGLKIIEVAGLVLPFLCHAIIGIVYLFQKSNCYSGDSRPHLRY
AKMYSYTMORWTAWILLFGIAFHVVHLRFIRYPVHVDIHGTTYYAVDIQPSRYDVIVRGT
KGFLTLNLPNTEASSIEVSRHDLGGADAALLSERNSYLLTPSAGTAFLYVVRDALGSLFI ALLATILVIAAAFHGFNGLWTFCCRWGVVVSLRMQGVLRIVCYLAMIVVTFMGVSAVWNL

CPH=0789 887436 889316
sdf#-Succinate Dehydrogenase
CMDENRKVIVVGGGLAGLSAAMQLANLGIIVELVSLTKVKRSHSVCAQGGINAALNLKPE
EEDSPYVHAYDTIKGGDFLADQPEVLEMCLAAPRIIKMLDNFGCPFNRGFSGKLDVRFFG EEDSPYVHAYDTIKGGDFLADQPPVLEMCLAAPRIIKMLDNFGCPFNRGPSGNLDVRRFG
GTLYHRTVFCGASTGQQLMYTLDEQVRRREHAGRVIKRENHEFVRLVTDHSGRACGIILM
NLFNNRLEILRGDAVIIATGGPGVIFKMSTNSTFCTGAANGRLFLQGMAYANPEFJQIHP
TAÐFGRDKLRLISESVRGEGGRVWVPGDSSKRIVFPDGSERPCGETGAPWYFLEDMYPAY
GNLYSRDVGARAILRVCEAGLGIDGRMEAYLDVTHLPEKTRIKLEVVLDIYKKFTGEDPN
TVÐFRIFFPAVHYSMGGAWVDMPAADDPDRDSRFRQMTNIPGCFNCGESDFQYHGANRLGA
NSLLSCLFAGLVSGDEASRFIEAFGASQATSSDFDRALQQEKEENARLLSASKKENIFVL
HEÐTAKIMYRNVTVKRNNRDLQETMDKLKEFRERLKNIVSVLDSSPFANKSFHFVRQMSPM LELALA ITKGALLRNEFRGSHYKPEF PERDDEHWLKTTVAVYAPEEPEI SYLPVDTRHVA PTËRDYTKSSTGKI ELTNI PDNI RLPI

CPn=0790 889279

CPHEDT90 893279 890103
sdbB-Succinate Dehydrogenase
NSRTFLIISVYPYRKREMMENLETFILKIYRGVPGKQYWESFELPLHPGENVISALMEIE
KRPVNILGEKVMPVWEQGCLEEVCGSCSILVMGVPRQACTALIQEYIDATQSREIVLAP
LTKFPLIRDLIVDRSIMFDNLERIQGWWAADIEGETFGPQVTQEQQELLYALSQCMTCGC
CTEACPQIDNKSDFIGPAAISQARYFNTYPGDKRSKKRWRALMGKGGIEGCGQAHNCVRV
CPKKLPLTESISAVGREISKFSLRSLFSALFKKKK

893104 890111

CPn_0791 893104 890111
CT590 hypothetical protein
TCLRSSRKIVVEDISDRMMYSCYSKGISHNYLLHPMSRLDIFVFDSLIANQDQNLLEEIF
CSEDTVLFKAYRTTALOSPLAAKNLNIARKVANYILADNGEIDTVKLVEAIHHLSQCTYP
LGPHRHNEAQDREHLLKMLKALKENPKLKESIKTLFVPSYSTIQNLIRHTLALNPQTILS
TIHVRQAALTALFTYLRQDVGSCFATAPAILIHQEYPERFLKDLNDLISSGKLSRIVNQR
ELAVPINLSGCIGELFKPLRILDLYPDPLVKLSSSPGLKKAFSAANLIETLODSEAQIQQ
LLSHQYLMOKLQNVHETLTANDIIKSTLLHYYQLQESTVRAIFFKEGLFSKEQVAFSTQH
PRELSEIQRVYHYLHAYEEAKSAFIHDTONPLLKAWEYTLATLADASQPTISNHIRLALG
WKSEDPHSLVSLVTHFVEEEVENIRILVQQCEQTYHEARSQLEYIEGRMRNPLNNODSQI
LTMOHMRFRQELNKALVENDSAQEKAKKFLHLEFFLLSFYTKQIFLYFRSSVDAFTQEFA
HLYANN PAGFRILFTHGKTHENTWSPIYSINEFIRFLSFFTSTESELLGKHAVINLEKE
TSÄLVHNITAMLHTDVFQEALLTRILEAYQLPVPPSILNHLDQLSQTFWVYVSGGTVDTL
LLDYFEGSEPLTLTEKHPENPHELAAFYADALKDLPTGIKSYLEEGSHSLLSSSPTHVFS
ILAGSPLFREAWDDWWSYTMLRDWWKCHQDPFLQDFILPGLSIYAFIENFCNKYALOHV
VHDFHDFCSDHSLTLPELYDKGSRFLSSLFTKDKTVALIYIRRLLYLMVREVPYVSEQQL
PEVLDNVGCYLGISGRITYEKFRSLIEETIPKMTLLGSADLRHIYKGLLMQGYQKIYTEE
UTYLRLTTAMRHINLAYPAPLLFADSMWESIYGFILNPGTTEIDLWKFNYAGLQQPLD
HIGGEFATGRFWTLYANPIDYGMPPPPGYRSRLPKEFF NIQELFATGREWTLYANEIDYGMPPPPGYRSRLPKEFF

394955 893108

CPD_0792 84455 893108
CT589 hypothetical procein
RHHLINIKOIRIMKHTETKRVLFEFFLVIFIFLLHLMVVOFFGEGAAKANLVQVLHTRA
THLGIEFEKKUTIHKLFLDRLANTLALKSYASPBAEPYAQAYNEMMALBYTDFSLCLIDF
FUGURTYNIMGDETRYLKQHEMKKKLBAAVGKAFLLTITGKPLLHYLILVEDVASWDS
TTTTCGLUVFFYENEDGKDLFGSLHTKKONLCUVHYGEVLFGAQDGEGFFESLDLPNL
FQFQARGEGAIEIEKANGILGENLITVSINKKRYLGLVLNIK IPTQFYTLBLUVPVSDLI
DGFQARGEGAIEIEKANGILGENLITVSINKKRYLGLVLNIK IPTQFYTLBLUVPVSDLI
DGALKVPLMICFFYVLAFLHMWNIFSKINTKLNKPLGELTFY MEAAWRCNHNVRFEPQPY

GYEFNELGNIFNGTLLLLLING TEKAD I DYHGGENLGKELGTU. JUGALLG PDFPTFPKV
TESSQHLRPRQLSGHFNGWTVQDXGDTLLG I IGLAGD I GUP 3YU YALGARGUFLAYASSD
VSLQKISKDTADSFCKTTEGHEAVVANTFIKYVEKDPSLELLGUSEGAPTMFLQRGESFV
RLPLETHQALQPGDRL ICUTGGED I LKYFSQL PIEELLKDPLNPLNTENL I DSLTMMLNN
ETEHSADGTLTILSFS

895838 994919 CPn 0793 rbsU-sigma regulatory family protein-PP2C phosphatase (RsbW

ETHER TO . 11:1 GENERAL SALE COME TILBERNAACKA ATTOTYPENVOLULES DVLDLDAGIFETPNVLLSNEMCKVFYGIYNEISLIKVFPNGA KIVVASSIPEHLGENYNHKIDIPKNTPFLAALKOSPKNGEVFSVMQANVFDAKTQELQGI LYTTFSAESLLKDLLINKQSYLTVKTAILSKYGVILKASDPALHLHTVYPDMTKEKFCQV FLNDDPCPIDSELGPLTLSPLDIGENFYSFKIKDTPIEWGCIENNYSIDIAVLSYAKKES FAPLWRRARMTTAYFFCILLGSLIAFIVARRLSLPIRKLATAMIESRKNKNCLYTDDSLG FEINRLGHIFNAMVENLHKQOHLAKTNFEMKENAQNALHLGEQAQQRLLPNTLPSYPHIE LAKAYIPAITVGGDFFDVFVVGEGSKARLFLIVADASGKGVNAGGYSLFLKNMLRTFLSR SSSLQQAIGETSRLFYNNTKNSGMYVTLCVVCYHQTSNTMEYYSGGHPACYLDPDGETS WLFHPGMALGFLPEVANITSKLFHPKPGSLFVLYSDGITEAHNNNDMFGEERLQAAIQG LTGKSAADAVHRLMLSVKTFVGNSHQHDDITLLILKVLES

897123 898004

CPH_U/94 \$99004

No robust homolog present in Genebank/EMBL as of 11/7/98

KSSKHRSFLLKKSGGNQVSLYQKWWNSQLKKSLCYSTVAALIFMIPSQESFADSLIDINL
GLDBSVECLSGDGAFSVGYFTKAGSTFVEYOFFKYDVSKKTFTILSVETANQSGYAYGIS

YDGTITVTOTCSLGAGKYNGAKWSALGTLTPLTGITGGTSHTEARAISKDTQVIEGFSYDA

SGQPKAVQWASGATTVTQLADISGGSRSSYAYAISDDGTIIVGSMESTITRKTTAVKWN NVPTYLGTLGGDASTGLYISGDGTVIVGAANTATVTNGNQESHAYMYKDNQMKD

CPn_0795 898008 899195
No robust homolog present in Genebank/EMBL as of 11/7/98
GTLGGANSSATGVSSDGSVIVGQAQTADKSVHAFQYYNGEMKDLGTLGGTSSTAKTVSPD
GKVIMGRSQIADGSWHAFMCHTDFSSNNVLFDLDNTYKTLRENGRQLASIFNLQNMMLQR

CPn_0796 899280 901340
No robust homolog present in Genebank/EMBL as of 11/7/98
SELYSSYLOPCLMMSTVRNSALPLECLSRSETFKKVRSHMKFMKVLTEWIYRKDLWYTAF
LLTAIPGSFAHTLVDIAGEPRHAAQATGVSGOSKIVIGMKVYDDDFFATTVGFGYIDCHLQ
PLEAVRPQCSVYPNGITPDGTVIVGTNYAIGMGSVAVKWVNGKVSELPMLPDTLDSVASA
VSADGRVIGGNNINLGASVAVKWEDDVITQLPSLPDAMMACVNGISSDGSIIVGTMVDV
SWRNTAVQWIGDQLSVIGTLGGTTSVASAISTDGTVIVGGSENADSQTHAYAYMKSVMSD
IGTLGGFYSLAHAVSSDGSVIVGVSTNSEHRYHAFGYADCQMVDLGTLGGPESYAQGVSG
DGKVIVGRAQVPSGDWHAFLCPFQAPSPAPVHGGSTVVTSQNPRGMVDINATYSSLKNSQ
QLQRLLIQHSAKVESVSSGAPSFTSVKGAISKQSPAVQNDVQKGTFLSYRSQVHGNVQN
QQLLTGAFMDWKLASAPKCGFKVALHYGSQDALVERAALPYTEGLGSSVLSGFGGQVQG
RYDFNLGETVVLQPFMGIQVLHLSREGYSEKNVFFPVSVSVAYSAATSFMGAHVPASLS
PKMSTAATLGVERDLNSHIDEFKGSVSAMGNFVLENSTVSVLRPFASLAMYYDVRQQLV
TLSVVMNQQPLTGTLSLVSQSSYNLSF

901552 902694

CPn_0797 901552 902694
No robust homolog present in Genebank/EMBL as of 11/7/98
VLILTWINVLTKLGLMMSKKIKVLGHLTLCTLFRGVLCAAALSNIGYASTSQESPYQKSI
EDWKGYTFTDLELLSKEGWSEAHAVSGNGSRIVGASGAGQGSVTAVIWESHLIKHLGTLG
GEASSAEGISKDGEVVVGWSDTREGYTHAFVPDGRDMKDLSTLGATYSVARGVSGDGSII
VGVSATARGEDYGWQVGVKWEKGKIKQLKLLPQGLWSEANAISEDGTVIVGRGEISRNHI
VAVKNKNAVYSLGTLGGSVASAEAISANGKVIVGWSTTNNGETHAFMHKDETMHDLGTL
GGGFSVATGVSADGRAIVGFSAVKTGEIHAFYYAEGEMEDLTTLGGEEARVFDISSEGND IIGSIKTDAGAERAYLFHIHK

CPn 0798 902810 903856

Or TOUSE HOMOLOG PRESENT IN GENEBANK/EMBL AS OF 11/7/98

VVFEIIFVVRVPMKKTCCONYRSIGVVFSVVLFVLTTQTLFAGHFIDIGTSGLYSWARGV
SGDGRVVVGYEGGNAFKYVDGEKFLLEGLVPRSEALVFKASYDGSVIIGISDQDPSCRAV
KWYMGALVDLGIFSEGMQSFAEGVSSDCKTIVGCLYSDDTETNFAVKWDETGMVVLFNLP EDRHSCAWDASEDGSVIVGDAMGSEEIAKAVYWKDGEQHLLSNIPGAKRSSAHAVSKDGS FIVGEFISEENEVHAFVYHNGVIKDIGTLGGDYSVATGVSRDGKVIVGHSTRTDGEYRAF KYVDGRMIDLGTLGGSASFAFGVSDDGKTIVGKFETELGECHAFIYLDD

905001 903940

CPn_0799 905001 903940

No robust homolog present in Genebank/EMBL as of 11/7/98

KREENMAAIKQILESMLSQSSLMMVLFSLYSLSGYCYVITDKPEDDFHSSSAVKWDHWGK

TTLSRLSNKKASAKAVSGTGATTVGFIKDTWSRTYAVRWNYWGTKELPTSSWVKKSKATG
ISSDOSIIAGIVENELSQSFAVTWKNNEMYLLPSTWAVQSKAYGISSDOSVIVGSAKDAW
SRTFAVKWTGHEAQVLPVGWAVKSVANSVSANGSIIVGSVQDASGILYAVKWEGNTITHL
GTLCCYSAIAKAVSNNGKVIVGRSETYYGEVHAFCHKNCYMSDLGTLCGSYSAAKGVSAT
GKVIVGMSTTANGKLHAFKYVGGRMIDLGEYSWKEACAHAVSIDGEIIVGVQSE

906550 905249

eno-Enolase

eno-Enolase

RKEIKIMFEAVIADIQAREILDSRGYPTLHVK/TTSTGC//GEARVPSGASTGKKEALEFR

RKEIKIMFEAVIADIQAREILDSRGYPTLHVK/TTSTGC//GEARVPSGASTGKKEALEFR

DTDDPRYQOKGVLQAVKNVKEILFPLVKGGSVYEQSLIDSLMMDSDGSPNKETLGANAIL

GVSLATAHAAAATLRRPLYRYLGGCFACSLPCPMMNLIN/SMHADNGLEFQEFMIRPIGA
SSIKEAVMMGADVFHTLKKLLHERGLSTGVGDE/GFAPHILASNEEALELLLLAIEKAGFT

PGKLISLALDCAASSPYNVKTGTVDGRHYEEG LAILSNLJORRYFIDS IEGGLAEEDVDGW

ALUTEVLGEKVQIWEDPLFVTNPELILEGISPGLANSVLIKPNQIGTURETVYAIKLAQM

AGYTTIINIRGGTTTMTIADLAVAFNAGQIKTGSLSPCEPVAKYNBIMEIEEELGSEAI PTDCMVP3YEDGCP

CPn_0801 908709 906727

IVEP EXTRICTOR OF ARC SUBURIT B
LEFTMTPCLHAFTANDAPEA LAPESGOVRIO/VKSOVI L/TTN) ZIPTTTANDVANDIL
PELVIZIRIATTANDA QPERPEPPRIAVEYET ZYYDYYQADAY LAPEDDE LERGILLINDE
LDKLIFEGARRET LERGITTLUS SOCCI TYG GOGHENYTSMALVLEWGEN, PROLLETAQUVK
MRYQA IPTORCAPRIBLING VOCCI TYG GOGHENYTSMALVLEWGEN, PROLLETAQUVK
MRYQA IPTORCAPRIBLING VOCCI TYG GOGHENYTSMALVLEWGEN, PROLLETAQUVK
DATLYPEGRYV I PEATREDA DET LGEGLLIFERMAFPDDP LEKDR FERRETHOLLHIKETG
PCKOTENYSRHPTOAPMARTCLED/PPEDFULLIFERGTOTLPQ (RAMYRGIZ) RRQSL

VEYGERLPGAFDNRPLTYEEAOKYFRKVIYVSATPGDTEVQESSGHIVGOIIRPTGIPDP WETTRATTOVODILEETRIPLISORHEKTLVISTTENDE SYNDER TUSTITETRITUSE ETAERTO LUTDLROGV LOVILGVNLLREGLDLPEVSLVATLDADKEGFLRSTSSLIGFOG RAARNINGKVIFYADOKTRS LEETLRETERREG LOLDVIKEHNIVEKP I KAI FANPILQ TSKDSESPRESORPLSKEDLEEGIKKYEALMORAAKEFRFNEAAKYRDAMQACKEGLLYL

CPn_0802 909761 908709
THE REPORT OF THE ANALYSIS FOR THE CONTROL OF THE CONTROL OF THE CONTROL OF THE CONTROL OF THE CONTROL OF THE CONTROL OF THE CONTROL OF THE CONTROL OF THE CONTROL OF THE CONTROL OF THE CONTROL OF T PSLKMARNASIEBGSLSYGLIGYPILQSADILLAKAÇFVPVGKDNEAHVELTRDIARNF NRLYQQVFPEPEVLQGELTSLVGIDOQGMSKSANNAIYLSDSDATITEKVRKMYTDPNR IRATTPGRVGGNELTYHDIFNPHKDEVEEFKARYRQGCIKDIEVKARLAEELIHFLKPI KERRSEFLSKPLALQNVLEDGTHKMREVAKVTMEEVHDKFGFSHKWRSLLK

CPn_0803 910306 909752
CT584 hypothetical protein
FMAAKTKTLELEDNVFLLLEENLKRIFATPIGYTTFREFQNVVFNCANGQQEIANFFFEM
LINCKLTOELAPQQKQAAHSLIAEFMMPIRVAKDIHERGEFINFITSDMLTQQERCIFIN
RLARVDGQEFLLMTDVQNTCHLIRHLLARLLEAQKNPVGEKNLQEIQEEITSLKNHFDEL

CPn_0804 911074 910310 gp6D-CHLTR Plasmid Paralog EIFSSMGNLKTLLESRFKKNTPTKMEALARKRMEGDPSPLAVRLSNPTLSSKEKEQLRHL LQHYNFREQIEEPDLTQLCTLSAEVKQTHHQSVLLHGERITKVRDLLKSYREGAFSSWLL LTYGNROTPYNFLVYYELFTLLPEPLKIEMEKMPROAVYTLASROSPOEKKEEIIRNYRG ERKSELLDRIRKEFPLVETDCRKTSPVKQALAMLTKGSQILTKCTSLSSDEQIILEKLIK KLEKVKSNLFPDTKV

CPn_0805 911846 911067
minD-chromosome partitioning ATPase—CHLTR plasmid protein GP5D
GYASRMKTIAVNSFKGGTAKTSTTLHLGAALAQYHQARVLLIDFDAQANLTSGLGLDPX
VDSLAVVLQGEKEIQEVIRPIQDTQLDLIPADTWLERIEVSGNLAADRYSHERLKYVLGS
VQXYDYVIIDTPPSLCWLTESALIAADYALICATPEFYSVKGLERLAGFIQGISARHPL
TILGVALSFWNCRGKNNSAFAELIHKTFFGKLLNTKIRRDITVSEAAIHGKPVFATSPSA RASEDYFNLTKELLILLRDI

CPn_0806 913816 911867

threonyl trna Synthetase NAMNESPPNMEAWNKMIQVTCDQKNYEVLEGTTAAELAKQLKNSHQFIGVLINERPRDLS NAMES PPMMEAMMKMIOVTCDOKNYEVLEGTTAAELAKQLKNSHOFIGVLINERPROLS
TELREGDTLVFLTSEDPEGREIFLHTSAHLLAQAVLRLWPDAIPTIGPVIDHGFYYDFAN
LSISSDFFLIEDTVKQIVDEKLAISRFTYGDKQQALAQFFONPFKTELIRELPENBEIS
A¥ŞQGEFFDLCRGPHLPSTAHVKAFKVLRTSAAVWRGDPSRESLVRIYGTSFPTSKELRA
HLEQIEEAKKRDHRVLGAKLDLFSQQESSFGMPFFPRBGMIYWDALIRYWKQLHTAAGYK
EEEFPOLMWRQLWEVSGHWDNYKAMMYTLQIDDEDYAIKFWMCPGCMLYYKTRLHSYKEF
PERVAEVGHVHRQEASGALSGLMRVRAFHQDDAHVFLTPEQVEEETLNILQLVSTLYGTF
GIEYHLELSTRPEKDTIGDDSLWELATDALMRALVQSGTPFIVRFGEGAFYGPKIDHVK DAIORTWOCGTIQLDMFLPERFELEYTTAGGTKSVPVMLHRALFGSIERFLGILIENFKG RFPLWLSPEQVRIITVADRHIPRAKELEEAWKRLGLVVTLDDSSESVSKKIRNAONMOVN YMITLGDHEINENVLAVRTRDNRVINDVSVERFLNTILEEKNSLSLTALL

CELEO807 913950 914879
CT580 hypothetical protein
TLQTGLHMSLFLVFLTAFIWSSSFALSKLVMNASAPIFATGARMVIAGAILALAAWFRGG FEGISKKIFLYIVLLALTGFYLTNIFEFIGLOSLSSKTCFIYGLSPLMSALFSYIOLKE KWFFEKKVLGLSLGLVSYICYLTFGGGGDSOPWTWQTGLPELLILGASLASFGWTLLRQ LEKOSTLSVTAINAYAMLIAGMLSIMHSAVVEPWRPLPVQDISQFLYATLALVVISNILC VNLTAKLLRKYSSTFLSFCNLVMPLYSGFYGWILLGEKGVSLGLVLAVAFMVAGCRLIYH EEFROGYIVS

CTPLUB08 916398 914956
CTPST9 hypothetical protein
LKKEFSWALKSLKRMPQSAEPSLAHIKPIIFKGACIAMTSGVSGSSSODPTLAAQLAQSS
LKKEFSWALKSLKRMPQSAEPSLAHIKPIIFKGACIAMTSGVSGSSSODPTLAAQLAQSS
CKAGNAQSGHDTKNVTKQGAQAEVAAGGFEDLIQDASAQSTGKKEATSSTTKSSKGEKSE
KSGKSKSSTSVASASETATAQAVQGPKGLRQNNYDSPSLPTPEAQTINGIVLKKGMGTLA
LLGTWMTLMANAAGESWKASFQSONQAIRSQVESAPAIGEAIKRQANHQASATEAQAKQS
LISGIVNIVGFTVSVGAGIFSAAKGATSALKSASFAKETGASAAGGAASKALTSASSSVQ
QTMASTAKAATTAASSAGSAATKAAANLTDDMAAAASKMASDGASKASGGLFGEVLNKPN
WSEKVSRGMNVVKTQGARVASFAGNALSSSMQMSQLMHGLTAAVEGLSAGQTGIEVAHHQ
RLAGQAEAQAEVLKQMSSVYGQQAGQAGQLQEOAMQSFNTALQTLQNIADSQTQTTSAIF
N

917794 916307

CPD_0809 917794 916307
CT578 hypothetical protein
DTNMSISSSGPDNOKNIMSQVLTSTPQQVPQQDKLSGNETKQIQQTRQGKNTEMESDAT
IAGASGKDKTSSTTKTETAPQQGVAAGKESSESQKAGADTGVSGAAATTASNTATKIAMQ
TSIEEASKSMESTLESLQSUSSAAQMKEVEAVVVAALSGKSSGSAKLETPELPKPGVTPRS
EVIEIGLALAKAIQTLGEATKSALSNYASTQAQADQTNKLGLEKQAIKIDKEREEVQEMK
AAEQKSKDLECTMDTVNTVMIAVSVAITVISIVAAIFTCGAGLAGLAAGAAVGAAAAGGA
AGAAATTVATQITVQAVVQAVKQAVITAVRQAITAAIKAAVKSGIKAFIKTLVKAIAKA
ISKGISKVFAKGTQMIAKNPFKLSKVISSLTSKWTVGVGVVAAAPALGKGIMQMLSEM
QONVAQFQKEVGKLQAAADMISMFTQFWQQASKIASKQTGESNEMTQKATKLGAQILKAY
AAISGAIAGAHKTNNF

CPn_0810 918193 917825 CT577 hypothetical protein GEIMNKKPKKTKKAVQSKAAPVKRVPEESQEAAIQQLELAVSDLYKELPLAQTFASLTDK NQ INSI IAALSOTLESLHLEELTQSLFFSAQEDANFAKELSSVVHGLKNLTTVVNKQMVK

CPn_0811 018933 018208

CCH-LOW CO RESPONDE FROTOITH H
LONNETLATREMERSPERANDOPORESSAITHKETE SPLAELAAOKKAKADDLEDVHPVPT
ERETEKALAIM FERDI-INKLDIDG LOGISDVELLET (TVAYTEYSOGKYNEAYSLEGILAA
AQPONYKYMIA)LISCYHOLHLYNDAARGEFLAEDAG POND EPPYY LADSILLKIQOPEESN
NELDVIMDTCONNEGEKTILKERCOTMKOGTEKQMAGDTKKAPTYKDAGKSKTTTNKKOGK

CPn_0812

919081 920862

muti-DNA Mismatch Repair
Gilicwignitrapmstrrpiqlidpltingiaagevienuvuvvkeliensidagadei
eietiggggaliirdngggepertunggaagevienuvuvvkeliensidagadei
eietiggggepertunggdivscepcarquottvivnsleynvpvreporepoksm
gsbrigirklienriistanigwswissenheigiakqqufqevvvvmgdhfmqdalti
DKEANGVRIVGVLGSPSFHRPTRQQQKIFINDRPIESLFIXKKVDAYYLLDEHRYPVF
VLKLYLPSSWCDFNVHPQKIEARILKEELVODCIKARIVETLACPPGILGRTHQEIEESD
SVPLPMFRMLETSDVGEESVEFDQNLFAYSSEDVGLEKOFYTGRGPKSQMDWIYSSDVR
FIT GROVLAFD. CHIETIAAN FISCHIMIN M. A. A. E. L. STEDEA
LT BERGVLAFD. CHIETIAAN FISCHIMIN M. A. A. E. STEDEA
LT BERGVLAFD. CHIETIAAN FISCHIMIN M. A. A. E. STEDEA
LT BERGVLAFD. CHIETIAAN FISCHIMIN M. A. A. A. E. STEDEA
LT BERGVLAFD. CHIETIAAN FISCHIMIN M. A. A. A. A. A. METCHINGARIA. BANGTOTI NSEALT
AL METCHIGATESKHOWEDVSBVSKIKLDMIVGRPERGFDJARIRRLIIDSDFMEG ALMKETLTQATFSKHQHVFDVSWLKLLWGVGKPEKGFDGAR LRRL1LDSDFMEG

CPn_0813 920843 921934
pepP-Aminopeptidase P
TLILWKDNHMSHDRILRAQRALSEHNLDAILVEKSEDLAYFLHDEAIAGILLIGQQEVMF
FVYRMDKDLYSHIQRVPLTFLTQDVVADLSLVVQKQRYQKIGFDSASTVYHKFAQRQVLP
CLWEPLECFTERIRSIKSEEEIRRMQEAAALGSAGYDYVLTLLREGITEKEVVRQLRAFW
AEAGAGEGFFPPIIAFGEHSAFPHSIFTDRPLKKGGIVLIDIGVLLNGYCSDMTRMTALG
TPHPKLLESYPVVVEAQKRAMALCKEGVLWGDIDAEAVRVLREHHLDTYFIHGIGHGVGR HIHEYPCSPRGSQVKLESGMTITVEPGVYFPGIGGIRIEDTLCIDKNKNFSLTARPVISE

921996 923357

CPn_0814 921996 923357
CT814.1 hypothetical protein
FFLFFKLSYNFIFNLPLTMYQLLSIGYSFVSFIALLWMLCYSPNYVTDLYRISLSAEESL
GGIRAFPQAESLLGGACALNFPDLEERLPDLRKELLFLGSNDRFDACGGKFSLQLASSKE
CYIAALKERVYLNVTNSSROPVYSFSPKOVPPELWTECFSVSVDGRVEVKVRLQGLHKEL
ISKPRDCETLFLNPPANKLDCWEIAGFRVDASFPVKQKIRRIGVDKFLLMHGGAEYADKA
TKERVDFVSSDEENYSRYLAVGDVLLWDGNCWQTCGEFQCASSRAPLFEVKRIDDKVMIA
DLWNVGGTORQTISLVKGVPSP IE INEVIREIEFTGMRSWKFPLVLVGQRLILSPDDWV
LRTAKGWEKLSRADQIQDYVTGKVTGPLLVFEKLEKDLRGFVLRGHMFNAQRTLVETISL
PLKQGFEPAVASQEVSSNTRSAAAHPGATNRGGS

CPn_0815 923361 925622
gspD/pilQ-Gen. Secretion Protein D
mVFFRNSLLHLVALSGMLCCSSGVALTIAEKMASLEHSGRGADDYEGMASFNANMREYSL
QLSKLYEEARKLRASGTEDEALWKDLIRRIGEVRGYLREIEELWAAEIREKGGMLEDYAL
WNHPETTIYNLVTDYGTEDSIYLIPQEIGAIKIATLSKFVVPKESFEDCLTQILSRLGIG WNHPETTIYNLVTDYGTEDSIYLIPQEIGAIKIATLSKFVVPKESFEDCLTQILSRLGIG
VRQVNSWIKELYMMRKEGCSVAGVFSSRKDLEALPETAYIGFVLNSNVDAHTNQHVLKKF
INPETTHYDVIAGRWIFGSAGEVGELLKIYNFVQSESIRQEYRVIPLTKIDRGEMISIL
NAAFREDLTKDVSEESLGLRVVPLQYQGRSLFLSGTAALVQQALTLIRELEEGIENFTDK
TYFWNVKHSDPQELAALLSQVHDVFSGENKASVGAADCGSQLNASIQIDTTVSSSAKD
GSVKYGNFIADSKTGTLIMVVEKEVLPRIQMLLKKLDVPKMVRIEVLLFERKLAHEQKS
GLNLLRLGEEVCKKGCSPSVSWAGGTGILEFLFKGSTGSSIVPGYDLAYQFLMAQEDVRI
NASPSVYTMNQTPARIAVVDEMSIAVSSGKDKAQYYRAQYGIMIMMLPVINVGEDGKSY
ITLETDITFDTTGKNHDDRPDVTRRNITNKVRIADGETVIIGGLRCKQMSDSHDGIPFLG
DIFGIGKLFGMSSTSDSLTEMFVFITFRILEMPVEQQERKEEALLSSRPGEREEYYQALA
ASEAARAAHKKLEMPPASGVSLSOVERGEYDXCC ASEAAARAAHKKLEMFPASGVSLSQVERQEYDGC

CPn_0816 925600 927102
gspE-Gen. Secretion Protein E
RGKNTMASILSQELLDILPYTFLKKHCLLPIEESSEAITIAHATATSVIAQDEVKLLIK
KPVRFVLKEESEILGRLQQLYSNREGNVSDMLLTMKEEDGTTISEEEDLLETTDTIFVVR
LLNNILKEAIEERASDIHFEPCEDSMRIRYRIDGVLHDRHSPPSHLRSALTTRLKVLAXM
DIABHRLPODGRIKHHIGGGEVDMRVSTVPVIYGERVVLRILDKRNVILDIAGLHMPKGT
EILFKDTITAPEGILLVTGPTGSGKTTTLYSVLQELKGPLTNIMTIEDPPEYKLPGIAQI AVKPKIGLTFARGLRHLLRQDPDILMVGEIRDQETAEIAIQAALTGHLVVSTLHTNDAIS AIPRLLDMGIESYLLSATLVGVVAQRLVRTICPYCKVAYTPENQEKSFLASLGKDTEMPL YRGQGCVHCFRSGYKGRQGIYEFLRPNTLFRSEVASNRPYHILRETAEQNGFLPILEHGI ALAVSGETTLAEVLRVTKRCD

CPn_0817 927106 928 gspF-Gen. Secretion Protein F

gspf-Gen. Secretion Protein F
GGRMPRYRYTYLDPKERRKRGYLEALHIQEAREKLAQENIQVLDIREVALRRMSIKSTEL
IVFTKQLLLLLRSGLFLYESLVSLRDQYHEQKMGLLLTSFMETLRSGGSLSQAMAAHPNI
FDHFYCSGVAAGESVGNLEGCLQNIIVVLEERAQITKKMVGALSYPCVLLVFSFAVMLFF
LLGVIPSLKETFENMEVKGLTKIVFGVSDCLSAYRYLFLGFASALITVGILMRHRIFMKK
LLEKLLFALPGTKKFVVKVAVNRFCSVASAILKGGGTLIEGLDLGCDAIPYDRLKTDMRD
IVQAVIGGGSLSQELAQRSWVPKLAIGMIALGEESGDLADVLGYVAHIYNEDTQKTLASI
TSWCQPVILIFLGGLIGVIMLAILIPLTSNIQTL

CPn_0818 928158 928682
predicted OMP [leader (16) peptide]
GYTKNVGFDNVVVSTRDSDFSWWPDRCDHVGNIDPTHKQYPNIIKCVLRGVGMKRQKRKQ

SITLIEMMVVITLIGIIGGALAFNMRGSIHKGKVFQSEQNCAKVYDILMMEYATGGSSLK EIIAHKETVVEEASWCKEGRKLLKDAWGEDLIVQLNDKGDDLVIFSKRVQSSNKK

CPn_0819 929117 928956 CT568 hyporhetical protein ASLYCYCLFLIWEKFHNNIGKANFHLKIITTDFLTDIYIVTIRDPIAYPLTGIC

929042 929659

CFICOSCO 122042 92959
CT557 hypothetical protein
DESLPCRCCCOTFPRSETSIRTEMPMCNSIAMKKOKRGFVLMELLMSFTLIALLLGTLG
FWYRKIYTVOKOKERIYNFYIEESRAYKOLRTLFOMSLSSSYEEPGSLFSLIFDRGVYRD
PKLAGAVRASLHHDTKDORLELRICNIKDOSYFETORLLGHVTHVVLSFQRNPDPEKLPE TIALTITREPKAYPPRTLTYQFAVGK

929637 930668

CEN_0821 92% 37 930668
CTS-56 hypotheticul protein
HTNLPLORKEMOPP LETLICHTSLVOLVAFDAANARKPCACAQTIERGENIFSIKRSACA
ELEVOEKSRHANALERISKOKGKVPPYGIAKVATKKKQRYRLLQVPISRPHNIGRYNLYA
LL. TIPECYSDTASWYALFIRLLPRAYVDTSHVPYSBEYALANALISRKQELLEPAQLIG
DWITTETEPERQAELFYKMEKGOSHOSOLIANFLHVPERGOGHEKLINLIFMDELLEPAVL
DHEDAYRETISLIRDI WCAVKPQEHALQEROQAAALELFYTRTDERLELRDKMOLLLSRY
DULELLRKKMEDYTLSDAGDYLFLVDFDYKALDERROTEKL

CENCERTS 430440 (1000) 1312/0
CENCERTS AND CENTRAL OF THE CONTROLL OF THE CONT

CPn 0823 932424 231501

CPH_D823
yscT/spa#-YopT Tranlogation T
FYALQVRFSKTSINGNKELMGISLPELFSNLGSAYLDYIFQHPPAYVWSVFLLLLARLLP
IFAVAPFLGAKLFPSPIKIGISLSWLAIIFPKVLADTQITNYMDNNLFYVLLVKEMIIGI
VIGFVLAFPFYAAQSAGSFITNQQGIQGLEGATCLISIEQTSPHGILYHYFVTIIFWLVG
OHRIVISLLLQTLEVIPIHSFFPAEMMSLSAPIWITNIKMCQLCLVMTIQLSAPAALAML
MINLFELBIHBMAEUVLVISTALIAGEASHMILLTTI WWWTIKQULVTTLAWFKLVPIML

CPn_0824 932677 932378
ysc5/f1iQ-YopS/f1iQ Translocation Protein
IRTRAVLAFFATSFKSVLFEYSYOSLLLILIVSAPPIILASIVGIMVAIFQAATQIQEQT
FAFAVKLVVIFGTLMISGGWLSNMILRFAGQIFQNFYKWK

933618

CPH_U825 935618 932677
yscR-Yop Translocation R
ERIKVFTIMRSIFRFSLCFFTLSVSCCFADASLYENSCPSRCQPTPPPSNSNPLNVVQQP
VAASSVPSYMPPLNADDVLPRDHLSDGSFSDTYPDITTQAILIFLLIFLALSFFLVMLLTSYL
KIIITLVLLRNALGVQQTPPSQVLNSIALILSTYVMFPTGVAMYXDARKEIEANTIPQSL
FTAEGAETVFVALNKSKEPLRSFLIRNTPKAQIQSFYKISQKTFPSEIRAHLTASDFVII IPAFIMGQIKNAFEIGVLIYLPFFVIDLVTANVLVAMQMMMLSPLSISLPLKLLLIVMVDGWTLLLQGLMISFK

934382 933612

CPT_U826

YSCL-YOP Translocation L

HDNKRSGVFSSEVNQPQRYYAIVKMKFFSLIFKDDDVSPNKKVLSPEAFSAFLDAKELLE
KTKADSEAYVAETEQKCAQIRQEAKDQGFKEGSSWSKQIAFLEEETKNLRIRVREALVP
KAIASVRKIGKELELHPETIVSIISQALKELTQNKHIISVNPKDLPLVEKSRPELKNI
VEYADSLILTAKPDVTPGGCIIETEAGIINAQLDVQLDALEKAFSTILKAKNPVDEPSET SSSTDSSSLSNDQDKKE

CPn_0827 935273 934434
CT560 hypothetical protein
GCLVTANTFGTLDILMKHSKEDDLSRFLPKNLLVESPHPEEIPLKSLSFTMSWLPTIHPS
WITIAMKEFPPEIQGQLLAWLPEPLVQEILPLLPGISIAPHRCAPFGAFYLLDMLSKKIR
PCGITEEIFLPASSANAILYYTGPVKIALINCIGLYSIAKELKHILDKVVIERVKNALSP
TEKLFLTYCQSHPMKHLETTNFLSSWTTDAELRQFVHKQGLEFLGKALTKENASFLWYFL
RRLDVGRAYIVEQTLKTWYDHPYVDYFKSRLEQCMKVLVK

CPn 0828 936292 935267

yscilyop Translocation J IKRYAWIMVRRSISFCLFFLMTLLCCTSCNSRSLIVHGLPGREANEIVVLLVSKGVAAQK LPGAAAATAGAATEOMMDIAVPSAQITEALAILINQACLPPKKGTSLLDLFAKQCLVPSEL
DEMERYQEGLSEOMASTIRKMEGVVDASVQISFTTENEDNLPLTASVYIKHRGVLDNENS
IMYSKIKAKLIASAVPGLVPENVSVVSDRAAYSDITINSPWILTEELDVSSWGILIAKSS
LTKKRLIFYVLILILFVISCGLLWVIWKTHTLIMTMGGTKGFFNPTPYTKNALEAKKAEG AAADKEKKEDADSQGESKNAETSDKDSSDKDAPEGSNEIEGA

936729 937298 OPPLIANCE THE PROPERTY OF THE PVWFLFKCYS

CPB_0830 937339 937959

NOFFOBUST homolog present in Genebank/EMBL as of 11/7/98

DSCSFLLECFEVEAOTFPOVPSKVVVYKYKSSRILLIALLYNITLVIGGIFIHKKYLGOK
GRVILKIYONEEFFFATERFPSIGAGYLRVRNKNSVLFPFEDLMLVCPSVPKDFPLSAF
KVTTKLIYWSVLESIPVVGAFFFSIGRLFAMWCIEDFPGSIFSRIYHTIVGVLGILGIGI
IMFILRIIFTLLTLPFWLISCLKSSAA

CPn_0831 938249 938434
No rebust homolog present in Genebank/EMBL as of 11/7/98
NKREDNVLIRKSESEGAFFEATQNYPTIQQGYQLVRIREHNLSVRAHFDLSLSLDASVHP

939750 938827

CPH_0832 939750 938827
lipA-Lipoate Synthetase
VMKCRPTLNTDQFRVRKKLPERFPKWLQRPLPQGSAFHATDATIKRSGMPTVCEEALCPN
RAECWSRKTATYLALGDVCTRSCGFCNIGHSKTPPALDPTEPERIALSAKELGLKHVVIT
WVARDDLEDGGAQGLVDIIQKLREELPQATTEVLASFOGNVSALHTLDSGTITYNHNV
ETVARLSPLVRHKATYARSMFMLEQAANYLPDLKIKSGIMVGLGEMEGEVKQTLQDLASI GVRIVTIGQYLRPSRKHLQVKSYVTPETFDYYRRVGEAMGLFVYAGPFVRSSFNADMILA

941171 939747

CPn_0833 941171 939747
lpdA-Lipoamide Dehydrogenase
RGVLFEILITVSENMTQEFDCVVIGAGPSGYVAAITAAQSKLRTALIEEDQAGGTCLNRG
RGVLFEILITVSENMTQEFDCVVIGAGPSGYVAAITAAQSKLRTALIEEDQAGGTCLNRG
CIPSKALIAGANVVSHIKHABQFGIHVDGYTIDYPAMAKRKNTVVOGIRQGLEGLIRSNK
LTVLKGTGSLVSSTEVKVIGQDTTIKANHILLATGSEPRPFPGVPFSSRILSSTGILEL
EVLPKKLAIIGGGVIGCEFASLFHTLGVEITVIEALDHILAVNNKEVSGTVTNNKFTKQGI
RILTKASISAIEESQNQVRITVNDQVEEFDYVLVAIGRQFNTASIGLDNAGVIRDDRGVI
PVDETMRTNVPNIYAIGDITGKWLLAHVASHGOVIAAKNISGHEVMDYSAIFSVIFTHP
EIAMVGLSLQEAEQQNLPAKLTKFPFKAIGKAVALGASDGFAAIVSHEITQQILGAYVIG
PHASSLIGEMTLAIRNELTLFCIYETVHAHPTLSEVWAEGALLATNHPLHFPPKS

CPH_0834 141544 942014 CPT56 hypothetical proferm RIEMPFAKETEMORTCWCCCCVCMHVPCCPYCCNFLQDPPVACKOFSSCHICFPEGASK REAREDEFAVSGEWHAVLSENNPTOETNKOVIFEMTWLOCMPLAALFLGIGLLAFAFLIL LECTLOGILVLTWPKNRAYFYGIIGAAVAYRGYRKLEL

CED_0835 945698 942045
mot L SWIZERF Lamily helicase
ENTEMPLEALSTERODAMOLLKHERETVODOGEDSYFIKTEDDERSEGYWLOTIKLODID
RITTEASCEINIKESCHIMMINYFAVYDALSTHEELHEZPHESEWYAVDSHEFLDSIPLOAQ
RITTEASCEINIKESCHIMMINYFAVYDALSTHESELFTVOTNKTELKSALYPTAKKEFFL GEMYZTLESTHUTUTESTLSEEVFODWURTHASEELTVETNKTEURSALZPTAKKFFFL NEBSAEUTIGENSQSFFSHFSLOWQGLVPKAFTUDFFZTLEDTFFKLGLAHTGLENVSHDI

SITNATVOCAEEAKVNFTLSPVIHKKDRENHPKTRIGGVEYVAKTHEMITGFKALALPIYA
IPLLADKFKDQLLSLLCYDSLEYRCRYDIRLLRDASFSFSAYLVTPGDLDNGSLIYPNYC
YSPTKGLMQVVGMLSPKQAFIVKSEQVEDFLAERGHLIQEFGFGTFINERPEGHLTYNVT
EQGVLLFHYDVGDPSSTEIRFGTWTYYTNQGFFLEKKNDLPIQDGLIVEPQDIPAFIVKN
DAALRRLPNFFSSPPNLKDLLIEVHRQSRGKGLDLKPILVGLGESRCWLFGVFLYREDIG
FSLIETPLQGLCFLPRVIPEDVPQFLTQYAQHERILFPNPGTRPESYELVIQSIHRPH
PASPLHLQLEKTNLGSVPIGIALQGLKSKHTFLFTQAGFLDLKQNLFQFLKQFLSTQKC
VIAEMTVIANITDVFKLDALAPLSVTDTTIANPEDLOFFSQLKAACLPPIPONLFSSDHQ
LOG SOLDMANG LOMBEL SOLDE SOLDE SOLDE STOLD SOLDEN SOLDE SOL VFDEHMARNSQIRKIERTBAGGETETFINNEEFROEDETFINNEERROEDETFINNE PVDAPVEDDTGVLPPEDS

CPn_0836 946960 945722
brnQ-Amino Acid (Branched) Transport
KMKKNASHKTNDKKSLSIWSIGGSIFAMFFGAGNIVFPLALGYHYNAHPWSÄYFGMMLTA
VCVPLLGUVSMLFYSGDYQKFFFSIGRIFGMIFITAIILLIGPFGGIFRAIAVSHATLIS
LSEHKSAFIPSLPIFSAICCVLIYIFSCKLSRLIQWLGSVFFPIMLVTLLWVIIRSFMIP
THPMVQEFIPNARQAWLAGFIEGFNTMDLLAAFFFCSIVLISLRQLVAEEKHPTEEEIPL
SFQGISKNKRSLALGFFLAAILIGMTYLGFVLSAAFHAGLLVNVSKCHILGRISAIALG
PNSILAGVSVFIACLTTEIALVGIVADFLARVVSFKKLNYASAVICTLIPTYLISILNFE
TISHLLLPLLQLSYPALIVLACGNIAYKLWNFRYSPVLFYLTLSLTIVLKLVN

CPn 0837 947777 947145

TITLE TO THE TOTAL THE TOT YARQYCPALHHKIDNCPICSYLAKEANSTRT

CPn_0838 949196 947781

thdF-Thiophene/Furan Oxidation Protein
ISLNIYPNSFHLFNLKIGILSESSFNFSIFMLKHDTIAAIATPPGEGSIAVVRLSGPQAI
VIADRIFSGSVASFASHTIHLGQVIFEETLIDQALLLLMRSPRSFTGEDVVEFQCHGGFF
ACSQILDALIALGARPALPGEFSGRAFLINGKIDLVQAEAIQNLIVAENIDAFRIAQTHFQ
GNFSKKIQEIHTLIIEALAFLEVLADFPEEGPDLLVPQEKIQNALHIVEDFISSFDEGQ
RLAQGTSLILAGKPNVGKSSLLNALLQKNRAIVTHIFGTTRDILEEQWLLQGKRIRLLDT
AGQRTTDNDIEKEGIERALSAMEEADGILWYIDATQPLEDLPKILFTKRSFLLWNKADLT
PPPFLDTSLPQFAISAKTGEGLTQVKQALIQWMQKQEAGKTSKVFLVSSRHHMILQEVAR
CLKEADQNIYLDPBFILALEIPEBLHSIGME GKNEUTESILGFIESKEPIGK CLKEAQONLYLQPPEIIALELREALHSIGMLSGKEVTESILGEIFSKFCIGK

CPn_0839 949230 950159
psdD-Phosphatidylserine Decarboxylase
FLFIVSRGLVQKPQYIDRITKKKVIEPIFYEKTMLFLYNSKLGKKLSVFLSTHPIFSRIY
GWLQRCSWTRRQIRPFMNRYKISEKELTKPVADFTSFNDFFTRKLKPEARPIVGGKEVFI TPVDGRYLVYPNVSEFDKFIVKSKAFSLPKLLGDHELTKLYAHGSIVFARLAPFDYHRFH FPCDCLPQKTRCVNGALFSVHPLAVKDNFILFCENKRTVTVLETEQFGNVLYLEVGAMNV GSIVQTFSPNQTYAKGDEKGFFAFGGSTVILLFLPNAIRFDNDLLKNSRMGFETRCLMGQ SLGRSQREEI

CPn_0840 950141 951544
CT700 hypothetical protein
ISERRNLKTLKTFFGIAKROKSQKWRIMMLVILWALAASLAIALVAKGYYRFVYFRRYAV
QVIREVRLSMELKEWALAEQQLLPILKKRSYRRQCLFEYMRILRKMQRFEESEKLLAEAK
KLGLRGPYFFLEIAYKAYRFGAFKECAQAFASVPQDLFEEDAAKYASALVRLGDLDAAC
SLIEPWISPLSHQETFVTMGHIYFTSKRYKDAIDFYNRANALGVCPVEVTYNLAQAYRIT
SSYAKAGKLFRKLLSNPVYKEEALFNIGLCEQKLGRPGKALLIYQSSDLWSRGDALLMY
AAMAAMDQRDYVLAEPCWELALRCSTFAKDYKCGLGYGFSLCRLRKYGDAERVYCNLIQN
FPECLTACKALAWLCGVGYATLLGSEEGLMYAKKAVELDHSCETLELLSACEARCGNFDA
AYEIQSFLSSRDTSLQEKQRRSQILRILRKKLPLNDHHIVEVDALLAA

951719 954640

secA-Translocase SecA IKRHMLGFLKRFFGSSQERILKKFQKLVDKVNIYDEMLTPLSDDELRNKTAELKORYONG IKRHMLGFLKRFFGSSQERILKKFQKLVDKVNIYDEMLTPLSDDELRNKTAELKQRYQNG ESLDSMLPEAYGVKNVCRRLAGTPVEVSGYHQRWDMVPYDVQILGAIAMHKGFITEMOT GEGKTLTAVMPLYLNALTGKEVHLUTVNDYLAQRDCEWVGSVLRWLGLTTGVLVSGTLLE KRKKIYQCDVVYGTASEFGFDYLRDNSIATRLEEQVGRGYYFAIIDEVDSILIDEARTPL IISGFGEKHNPVYFELKEKVASLVYLQKELGSRIALEARRGLDSFLDVDILPKDKKVLEG ISEFCRSLALVSKGMPLNRVLRRVREHPDLRAMIDXWDVYYHAEQNKEESLERLSELYII VDEHNNDFELTDKGMQQWVEYAGGSTEEFVMMDMGHEYALIENDETLSPADKINKKIAIS EEDTLPKARAHGLRQLLRAQLLMERDVDYIVRDDQIVIIDEHTGRPQPGRRFSEGLHQAI EAKEHVTIKKESGTLATVTLQNFFRLYKLAGMTGTAITESREFKEIVLYVLQVPTFKP CLRIDHNDEFYMTEREKYHAIVNEIATIHGKONPILVGTESVEVSEKLSRILRQNRIEHT VLNAKNHAQEAEIIAGAGKLGAVTVATTMAGRGTDIKLDNEAVIVGGLHVIGTTRHRQSRR LDRQLFRGCARLGDGAAKFFLSFEDRLMRLFASPKLNTLIRHRPPEPECAMSDPMFNRL IETAQKRVEGRNYTIRKHTLEYDDVMKKORQAI*AFRHDVLHAESVFDLAKEILCHVSLM VASLVMSDRQFRGWTLPNLEEWITSSFPIALNIELRQLKDTDSIAEKIAAELIQEFQVR DPHLVEGSKAGGEELDLASAICRDVVRSVMVMHIDEQWRIHLVDMDLLRSEVGLRTVGQK DPLLEFKHESFLLFESLIRDIRITTIARHLFRLELTVEPNPRVNNVI PTVATSFHNVNYG DPLLEFKHESFLLFESLIRDIRITIARHLFRLELTVEPNPRVNNVIPTVATSFHNNVNYG PLELTY/TDSEDQD

CPn_0842 955015 954710 CT702 hypothetical piotein (frame-shift with 0843) KYYTPPTISRSWSNIALKTISEPEYDONOLLKTOSLLTTHVDTLLNAPKDFPNSKNOKH ILFCIALMITLDHYAQFLIAGNRRKFWIPYYNDOVWSEWTPFI

CPn_984* **8973** 955230 954994 CT762 hypotherical piotein (frame shift with 0843) NKNPLITUWRI KYMNYDDYDTPDTDHUGITEDLLDNSEEAASLDKYQETGVYVEENTDQC DELIVED OF KMT THORO

OLEELS VORHEESEAALPONTFPDFSEVFTEGFSPEEPCTIPESPOOAPKTLKIALIGRP NVGKOS I INGLLNEERC I IDNTPOTTPON ID ILYSHKORQVLF IDTAGLERKKKNSIE WISSSRTEKA ISRADICILLY IDATOKLOSYERRILG ISKRKKPHI ILINKWOLLEEVRM BHYCKOLRATOPYLOSKMLC ISATTKRILKK IFOA IDELHHVSNKVPTP IVNKTLASA LHRNHPQVIQGRRLR IYYA IQKTTTPLQFILF INAKSLLTKHYEYYLKNTLKSSFNLYGI PFDLEFKEKPKRHN

CPn_0R45 958151 255850

THE COMMENT OF THE CO DRIIFSSMREDALRRDFTVNOMYYDPFEDKVFDFVECTRDIEKKVIRAIGHPRLRFSEDK LRILRAIRFSSSLGFTLDPTTERAIIKEAPALVNSVSPERIMQELKKMLKRQPYGALSLL LKLKVLIFIFPELRDIPYSLLRTTIEFARKFNPTHFPEILFLDLFGGVSEEAATVAFGR LRISNKELKLIESMYEALPHFONOSGNRVFWAHFLASPTAPLFLELFSALQKDPSRQMF LSRVQELESRLEOFILRIKTSSPVVSAPDLIAKGISPGRLLGDLLREAEILSIENECLDK EKILLLLQEKGFWK

959383 958112

CIDX-CLP Protease ATPase
REHMNKKNLTICSFCGRSEKDVEKLIAGPSVYICDYCIKLCSGILDKKPSSTISSAPVSE TPSQPSDLRVLTPKEIKKHIDEYYIGQERAKKTIAVAVYNHYKRIRALLHNKQVSYGKSN VLLLGPTGSGKTLIAKTLAKILDVPFTIADATTLTEAGYVGEDVENIVLKLLQAADYDVA RAERGIIYIDEIDKIGRTTANVSITRDVSGEGVQQALLKIVEGTTANVPPKGGRKHPNQE YIRVNTENILFIVGGAFVNLDKIIAKRLGKTTIGFSDDQADLSQKTRDHLLAKVETEDLI AFGMIPEFVGRFNCIVNCEELSLDELVAILTEPTNAIVKQYMELFAEENVKLVFKKEALY AIAKKAKQAKTGARALGMILENLLRDLMFEIPSDPTVEAIHIQEDTIAENKAPIIIRRTP

CPn 0847 960019 959387

CPH_U847 959387

clpP-CLP Protease Subunit

KLFDEETCMTLVPYVVEDTCRGERAMDIYSRLLKDRIVMIGQEITEPLANTVIAQLLFLM
SEDPKKDIQIFINSPGGYITAGLAIYDTIRFLGCDVNTYCIGQAASMGALLLSAGTKGKR
HALPHSRMMIHQPSGGIIGTSADIQLQAAEILTLKKHLANILSECTGQPVEKIIEDSERD FFMGAEEAISYGLIDKVVTSAKETNKDTSST

CPn_0848 961556 960177
Lig/murl-Trigger Factor-peptidyl-prolyl isomerase
VQASSPAFPFKSNKKGCLVPRSLSNEQFSVDLEESPGCIVSALVKVSPEVLNKLNKQALK
KIKKEITLPGFRKGKAPDDVIASRYPTNVRKELGELVTQDAYHALSTVGDRRPLSPKAVR NSITOFDLQEGARVEFSYEAFPAISDLEWENLSLYQDEAASEISDSDIEKGLINIGMFF ATKEFVERPSQEGDFISISLHVSKSNDENASSAAIFENKYFKLSEEMTDAFKEKFLGIS TGHRVVETITSPEIQSFLRGDTLTFTVNAVIEVSIPEIDDEKARQLQAESLDDLKAKLRI QLEMQAKDKQLQKRFSEAEDALAMLVDFELPTSLLEERISLITREKLLNARLQYCSDEE LEKRKSELIKEAEEDATKALKLLFLTHKIFSDEKLTISREELQYMDVCSRERFGQQPPK DISNOTLQELVMSARDRLTYSKAIEHVLRKAELLASTPSA

CPRI0849 961752 965285
mocll/snf-swf/snf family helicase
AD#IHSYSRGEMLNFRKLRRDFSANILQDGKKLFEQGAVIDAKILSMNGETVCISAQVR
GLYDNIYECEIEVDRSESDTVDSNCDCSYNYDCQHIVALLFYLEQYFNEWAYARSADL
ETBREINEEVKKELKETFVAAATKEEERKDREHQKEILREVYHAANALSANPFFLPLEYL
EKDSAELAVLFVSVNEDTFAPANQPIEFQLVLRLPCRSKPFYISNIRTFLEGVLYQEPIV
LNGRFFFFTMQSFNASDRKLIDLLIRYVRYPNHTTEKKLKSAYLMPPALGVILAKMFEH
OLADBCGCSIGEFESFGLFCOR FEDELGWST TRAXWFUND FEDELGWST. LNÖRRFFFTMQSFNASDRKLIDLLIRYVRYPNHTTEEKLLKSAYLMPPALGVILAKMFEH
QLADRGGGSLGEKESFSGLFCGNLEEPLCWSLTPAKMKFNLDFFDMPYRALLMTPVILVD
DEVOPEOTMLLESDAPGITHHFVHRFSFGJKRAHLRSFSRLRDIA IPEALFGSFRENA
LPYEDEVAEIANVALLNSFVTLPYVDEVRAICDMSYLDGELEAKLHFLYGSLRVPAASLA
LQYQDVRAFISDEGILARNLVEERKMLEEVFSGFIYDERDGAFRVKSEKKIVEFMTETIP
ANGHRITFFNCPENLSGOFIYDETIFELSFREGSDINYYEADLKVMGLLKGVPLDLLUMDCI
SAKKRFLELPKAGQOSKGTRGKVNSGKLPCILVLDLEKIAPVVQIFNEIGFKVLDDLVQ
KCBLÖBSLTGISLDOFEALPVNFSMSERLIEIQKQIRGEIEFDFQDVPQQIQATLRSYQTE
GVHWLERLRKMHLNGILADDMGLGKTLQAIIAVTQSKLEKSSGCSLIVCPTSLVYNWKEE
FRKFRPEFFTLVILGVPSQRRXQLTALADRDVAITSYNLLQKDVELYKSFRFDYVVLDEA
HHIKHRTTRNAKSVKMIQSDHRILITGTTIENSLEELWSLFDFLMPGLLSSVDFYGKYI
RTGEMGNKADNMVALKKKVSPFILRRMKEDVLKDLPPVSEILYHCHLTESQKELYQSYA
ASAKGELSRLVKQEGFERIHHVLATLTRLKQICCHPAIFAKDAPEPGDSAXYDMLMDLL
SSLÄMSGHKTVVFSQYYRMLGIIKKDLESRGIPFVYLDGSTKNRLDLUNQFNEDPSLLVF
LISLKAGGTGLNLVGADTVIHYDMWMPAVENQATDRVHRIGQSRSVSSYKLVTLNTIEE
KILTLQNRKKSLVKKVINSDDEVVSKLTWEEVLELLQI

965254 966390

CPn_0850 965254 966390
mreB-Rod Shape Protein-Sugar Kinase
LGKKYWNCCRYDFMSPHRNLEKLKNFSNRLYNRALGRFDKVFNFFSGNVGIDLGTANTLV
YVRGRGIVLSEPSVVAVDAQTHAVLAVGHKAKAMLGKTPRKIMAVRPMKDGVIADFEIAE
GMLKALIKRVTPSRSVFRPRILIAVPSGITGVEKRAVEDSALHAGAQEVILIEEPMAAAI
GVDLPVHEPAASMIIDIGGGTTEIAIISLGGIVESRSLRIAGDEFDECIINTMRTYNLM
IGPRTAEEIKITIGSAYPLGDQELEMEVRGRDQVAGLPITKRINSVEIRECLAEPIQQII
ECVRLTLEKCPPELSADLVERGMVLAGGGALIKGLDKALSKNTGLSVITAPHPLLAVCLG
TCKALEHLTOPKKRKGNIV TGKALEHLDQFKKRKGNLV

CPn_0851 966378 968195
pckA-Phosphoenolpyruvate Carboxykinase
REFGIVMVMSTNIKHEGLKSWIDEVAKLTTPKDIRLCDGSDTEYDELCTLMESTGTMIRL
NPEFHENCELVRSSADDVARVEQFTFICTSTEAEAGPTNNWRDPQEMRELHOLFRGCMQ
GRTLYIVPFCMGPLDSPFSIVGVELTDSPYVVCSMKIMTRWGDDVLRSLGTSGKFLKCLH
SVGKRLSFGEADVSWPCNPKSMRIVHFQDDSSVMSFGSGYGGNALLGKKCVALRLASYMA
KOGGMLAEHMLIKGITNPEGKKKVKFSASFPSACGKTNLAMLMPKLPGWKIECIGDDIAWI
REGROGRLYAVVDEYGFFGVAPGTSERTNPNALATCRSNSIFTNVALTADGDVWRSGLTE
QPPEPLTDWLGKPWKFGGSPAAHENSRFTAPLRQCFSLDEEMISPQGVPLDAIIFGGRRS
ETIFLVYFALSMEHGVTIGAGMSSTTTAAIVGGLKLENDFFAMLPFCYNMAYYPGHWL
SPAENRGLKLPKTFGVMWFRKNNGGETLWPSFSENLFVLEWIFGRTDGLEDIAERTPIGY
LITHIQKFNLMGLNLDLQTVQELFGVDAEGWLAEVENIGEYLKIFGGDCFQQITDELLRIK
GELKEK

968274 970513

TTO ALL PROBLEM TO A CONTROL OF THE PROBLEM TO A CONTROL O

LNLSGNYROLTENMLPNTETSLPOEIIAQIRSFQNIVNOTIIASNTLLPTTMRLDTLIGV IYTYQCCATIFGMSYGTSTPAKONYIDAINQEKSYWQARANGFDVTSDQVFDQFATNIQS GTSYRGIDLFKNNKVNEINPIFLSQAASFLRYPYNLMSRSMYQTIEDAANRSITALDGLI SGWSTQIATFQTQKNSLDSLLKVFDTMKANKESFVTTAPLQMVYSSIMLDKYLPTQQNV IASLGIQMTYSNKAAKYLNELIKEITTFQSADIYYSLSIYUKQMNLQAVADPIGKAVGVL NDEKTRAMADITRCNKIKAAIDKMLVEIKADAELSKSQIRELVDTLTNFKSQSDDLIRN SCLLGFLSGLTLKAVNDPPATYEAFTAEIFTEPFNNWKRQLATFESFVIQGGOMSITPGG

470637 971803

CT712 hypothetical protein CT712 hypotnetica: protein nimhpkiekrnslpltavapvfeesyhpsvattvDyvdattlsrhltvlkdvikearnld lgkaflitsmkqofintotelaiiqasladqssresrkkeerifhqhlgkaapqaatatsg voptadpvadkmplqsafa/vlldkyipaqeealyalgrelnlsgyaqnlfsplldmiks fnsapintnlssyisgtsctanfaygyemilssrynnbvsqcrldiastvkakaalamsa svkanvslddagkkjeddiastyksldvihtqltdvmtnlasiftypglnkydpsyriv ggdlsiialqndekvlvdgkvdittavneggllnffttvltdvqnygdlaqtqqlmldle lkamqqqwslvsaslkllmcmyttvisgfkn

CPn_0854 972849 971806

ompB-Outer Membrane Protein B GPFDMNSKMLKHLRLATLSFSMFFGIVSSPAVYALGAGNPAAPVLPGVNPEQTGWCAFQL GFF LWASHALMAILKAIL ISSERF GIVYSESAH ITALAGAR PAR PLEGYMPET (TWCAFLOL CNSYDLFRALAGSLKFGF/GDYVFSESAH ITAVPVITSVTTSGTGTTPF ITSTTKANDFD LANSSISSSCVFATIALQETSPAAIPLLDIAFTARVOGLKQYYRLPLNAYRDFTSNPLNA ESEVTDGLIEVQSDYGIVWGLSLQKVLWKDGVSFVGVSADYRHGSSPINY I IYYNKANPE IYFDATDONLSYKEWSASIGISTYLNDYVLPYASVSIGNTSRKAPSDSFTELEKQFTNFK FKIRKITNFDRVNFCFGTTCCISNNFYYSVEGRWGYQRAINITSGLQF

CPn_0855 974001 972994

gpdA-Glycerol-3-P Dehydrogenase
GLMKQHIGYLGMGIWGFCLASLLANKGYPVVAWSRNPDLIKQLQEERRHPLAPNVVISPN
LSFTTDMKEAIHNAFMIVEGVTSAGIRPVAEQLKQITDLSVPFVITSKGIEQNTGLLLSE
IMLEVLGDSVTPYLGYLSGPSIAKEVLMGSPCSVVVSAYDSQTLKQIHEAFSLPTFRVYP NTDIKGAALGGALKNVIAIACGIAEGLSFGNNAKAGLVTRGLHENRKLAAIMDCKPETLN GLAGLGDLCVTCFSESSRNLRFGHLLAQGLTFEQAKAKIGMVVEGAYTALSAYQVAKHHK IDMPITTGIYRVLYENLDLKEGIALLLORNTKEEFL

975410 973995

CPH_U856 y73410 y73495
agX-1 Homolog-UDP-Glucose Pyrophosphorylase
GSRDRNVRLTVMTESVYSPSAMMVNSLADKLKAINQEHILDIWPSLSPKQQQRLFQQLTS
VDIDFFRKQQQLLSSPTAILKDFHPITSFASSGEDERAHAGTTLLKEKKVACVVLAGGQ
GSRLKCDGPKGLFPVSPIKKKPLFQLVAEKVRAASKLAGQPLPLAFMTSPLNTRQTRSFF
ESNDYFHLDPNQVDFFCQPLWPLLTLSGDLFLEDMDTLALGPNGNGCIATLLYTSGWEK
WKNAGIEMVSVIPIDNPLALPFDVELCGFHAMSNNEVTIKAALRQTAIEDVGILVKSHDS GKTSVIEYSEIPONERFALNEDGKLKYCLANIGLYCLSMDFIRHAYQOLPLYKVHKHAK QLGHTSLNEKNAWKFEEFIFDLFCYSDHCQTLVYPRQECFAPLKNLEGNHSPDTVRQALS DRERQLFHKVTGKKLSPNTTFELEADFYYPSTSTSLHWENKAFFEEPFFEAS

975808 975392

CT716 hypothetical protein
CT716 hypothetical protein
LLLLRQYIKTARGISRLMRDRLGSLSLILKVKIHKYLDTLHNQKRLALTVSRNIQATNKR
LADLHLERYEHFISRDNIKHYDILLEYLKTLQSSLYKQQSESLRFLEIHHQQLQELINRR
KITEKINNNKYSKDQEIGT

977115 975757

CPn_0858 977115 975757
f111-Flagellum-specific ATP Synthase
RNSETRNORRTRPSTFCFDSMNHLNKEKLHIHNWQPYRACGLLSKVSGNLIEVDGLSACL
GELCKISSTKDPNLLAEVIGFHNHTTLLMSLSPLHSVALGTEVLPLRRPPSLHLSDHLLG
GELCKISSTKDPNLLAEVIGFHNHTTLLMSLSPLHSVALGTEVLPLRRPPSLHLSDHLLG
GVLDAFGNPIDKKEDLPKTHRKPLLSLPPSPMMRQPIDQIFPTGIKAIDAFLTLGKGQRI
GVFSEPGSGKSSLLSAIALGSKSTINVIALIGERGREVREYIERHSNALKQORTIIIAP
AHETAPTKVIAGRAAMTIAEYFREOGHEVLFIMDSLSRWIAALGEVALARGETLSAHQYA
ASVFHNUSFTFERAGNDKGSITALYAILYYPKHPDIFTDYLKSLLDGHFPLTSQGKALA
SPPIDILSSLSRSAQALALPHHYAAAERLRSLLKVYNEALDIIHLGAYTPGQDEELDKAV
KLLPSIKAFLAQPLSSYCYLDNTLKQLEALADS

CPn_0859 977597 977055
CT718 hypothetical protein
VFLVTTPQSPGSLSQSHLPHPHDPWDTEPTSLPEDPNDKASQELHSLVHLFRKLSIHLLS
EVEKTVQQLKPDLLELALLICEKFLYKKLENPQELALLLSTALQRHTTLRSLTPIKVFLH
PEDLKTLTDWISTHELPMIKHAEFFPDTSCRRSGFKIETPNGILRQEISEELDHLLSVLT

CPn_0960 PS639 977608
fliF-Flagellar M-Ring Protein
RTLVFFONLAKKLTALGISFLGCLLICGVVSCAILFGRSSNPSLAPTQVKTEKTSGNWLK
LTQMGNPKLIESLTKKECLEKDLTSFHPIASAKVAIALSTEDDVMSPLHLSVILTLRKEE SLTPSLLFSITDYLCSSLFGLKREHISLSDNLGNLYIPESITVNSLFIHTLENYLGKIFP KEHFALAYHAKAEKPTLQLTLNENYIAHLTKEESEKIVAHTKHYLYQNYDDSYDIVIETL PFARLQNKKSFPAKVLIGSMILVISLMI'ALASFYLARHAYERVSPEPRKIKRGINISKL LEIIQKESPEKIALILSYLDPKKAEALLNRLPEDLKHQVLKYKL

CPn_0861 a=9752 978925
nifU-NifU-related piotein
ASYPPTWKRLMTLPLEPMIFWSSLSAK/MKKFLTPHCAGTFSEEDAEAKEAHLVTGKQGH
PLMGNCVTFYWLVDKKNGVILDAKFQYFGHPYLIPLAEAVCNLVCGKSYSEAYKMTLDDI
DKSLRVHAHQPALPEDSISLYHFVIDALDTAVEQCLEIPLEDGSLPLONSPMNLDFEDAN
PYSQSDWEALTHEOKLYALRATIAEKIOPYIAMCOGEVTVESLENFIVTIAYSGNCSGCP
SSLGSTLNSICQLLRAYI\FELOVKVDESSLNLCHP

CPN_0862 080/24 779702
yitho Nits-related diotetin
grotiepitocomik ismerfolirkappiewinnovalppservkesyalikodipseppo
alkilaekteedirgoloviskooliterepyhelityviiitylaalvenismegirniiiilpah
boolulinslorhoolotty isvytynheypiveegilietiipprolesisaaholotyviop
libelislokorriliiilotistiloraektpeitiinao Littessaaloomosiooliterksil
ervpsomepphesatiaiilotistiloraektpeitiinao Littessaaloomosiooliterksil
glabsevonreniyvaaledipaesilaehilooliypsidyerpoplagvion gispp
lichsalhesiterskolitytklaramikalikiitpelesse

CPn_0863 Q8/S20081520 080831 980831 pgmA Phosphoglycer it e Mutara

EHMALLILLPHGQSVWNEKNLFSGWVDIPLSQQGIEEAFSAGRAIQNLPIDCIFTSTLVR SLMTALLAMTNHHSKKIPYIVHEDPKAKEMSRIYGAEEENNMIPLYQSSALNERMYGELQ GKNKKQTAEQFGEERVKLWRRSYKTAPPQGESLYDTKQRTLPYFEKNILPQLQNGKNVFV SAHGNSLRSLIMDLEKLSEEEVLSLELPTGKPVVYQWKNHKIEKHPEFFG

981658 982374

IVVSEGKKHEIRLFADAAGFPILELKRIRIGSLVLGGLRYGEYRELTDAELGTYMKLSD

CPn_0865 982412 982942 CT865 hypothetical protein SPMGYVFYVIAGSIFLGISLGAYCQLYYSVKSVLFSWYLLTVYALEKRHALLALSQLVGE EDAQSQKEIDFLSQCDKLSWRAFLKNSYEIIPFFKEMEDLLSERVQGFLESIETIAEHDR AILCIENFWASKNLFDFEIAAYEEAVEKYLKLRQRAPLRLASKLFRFLDVPSIRFSS

983494 982916

bira-Biotin Synthetase

MKVIYYEIEEIPSTNYMAKSYMHLWDPYALTVISTKCOTAGTGKFGKSWKSSKGDLLNT
FCFFITDLHIDVSRLFRLGTEAVVALCKDLGITEAKIKWPNDVLVHGEKLCGVLPETLPV
EGLLGVVLGIGLMGNTTKQALKDVGOPATSLQEILGHPIDLETTRELLIHHLLGVLQENL

PDSLATKSNRGNI

CPn_0867 983405 98466/
rodA-Rod Shape Protein
CIRIPQMHIGFCHCVRGGNFFYFVINNFHILEIYSLLNSNTIMRYHKYFRYVNSWVFLVV
LTLMLLSVVVISSMDFTAMLVTSSKGLLTNKSIMQLRHFALGWVFFICAYFDYHLFKRW
LTLMLLSVVVISSMDFTAMLVTSSKGLLTNKSIMQRHFALGWVFFICAYFDYHLFKRW
AUGUSTUS AU AWILYFMICALVGLFFVPSVQNVHRWYRIPFIHMSVQPSEYGKLVIVIMLSYLLESRKA DITSKTTAFLACLVVALPFFLILKEPDLGTALVLCPVTLTIFYLSNVHSLLVKFCTVVAT IGIIGSLLIFSGIVSHQKVKPYALKVIKEYQYERLSPSNHHQRASLISIGLGGIRGRGWK TGEFAGRGWLPYGYTDSVFSALGEEFGLLGLLFTLGLFYCLICFGCRTVAVATDDFGKLL AAGITVYLAMHVLINISMMCGLLPITGVPLILISYGGSSVISTMASLGVLQSIYSHRFAK

CPI_0868 986733 984670

znta/cada-metal transport P-type ATPase
NFRNGLGVRDLHHFREYYLITNEIIITGRYVFSRLFFTSFSAEVVNTFFESGMSEDTSPL
LSKORRKLSHNLPLKSAYLSLGTYLIALLSFWLHAKNLSNLFVVPTFFLAGTPALIKSLD
NICOKVVNIDILMTSAAFGSIFIGGALEGALLLVLFAISEALGQMVSGKAKSTLVSLKQL
AFFEGMLVLEDGNLQKVAINKIEVCNILRIKSGEVVPLDGEILHGSSSINLMHLIGEKVP
KSGHFGSIVPAGAHNMEGSFDLRVLRTGSDSTIAHIINLVIQAQNSKPRLQQRLDKYSSV
YALSIFAIACGIALLVPLFTSIPLLGPQSAFYRALAFLIAASPCALIIAIPIAYLSAINA
CAMHGVLLKGGVILDRLVSCNSVVMDKTGTLTTGELTCIGCDYFGSKNETFFPSVLALEQ
SSSFFIAEAIVSVLMEQKVSSLPADRYLTVPGEGVKGYFNEQEAFVGKVETGLGKVPSEY
LEDITECKIVQAKQHGEICSLAYVGNSFALFYFDLIPRPQAKEIIQDLKDLGYPVSMLTGD
HKVSAENTAEILGISEVFFDLIPEDKLAKIRELATORQIMVGGGINDAPALQATVGIA
MGEGGSATAIEAADIVLHHDSLSSLPWIIQKAKQTKKVVSQNLALALAIILLVSWPASLG
IIDLMLAVILHEGSTVIVGLNALRLLKS

CP#r-0869 987479 986658

CTIZE hypothetical protein
EGMEFFFPKTSENTSDCROHOILRKIMTODPHDHFKSRTPEDHIKHVRDKHRVCKGEPHT
FFKGFFYKLANNALSTOVFIFFIRTLFFLIPTNRALOVKSLISLGVGWTFYHGCLKARKA
WARMELSHRSMLEEKNEIEENFEQEKIELRILFENQGFKDPLLQEMVETVCSDSTLLLDT MIREELYIRKEDLPHPLIOGGSRILGGLCGLAIFLPLVLCISYTLAGVFSALMVLVLSFL KARILKNDKISEMVWVLGIFITSASIISSLMKLL

CPn_0870 988881 9 setg-Seryl tRNA Synthetase-2

setä-setyl trna synthetase-2
TTHPHOGOFGGAVILPFSPISIARRIKKSCCSEKSSIYSHFCTLLLNNETSMLDIKIIRK
TPEEETRLRKKDPKISLEPVLSLDKEVRQLKTDSETLQAQRRLLSQDIHKAKTQGVDAT
NLIQEVETLAADLEKIEQHLDQKNAQLHELLSHLPNYPADDIPVSEDKAGNQVIKSVGDL
PIFSFPPKHHLELNQELDILDFGAAAKTTGSGVPAYNNFGVLLEWALITYMLOKQAAHGF
QLWWLPPLLVKKEILFGSQJIPKFDGQYYRVEDGEQYLYLIPTAEVVLNGFRSQDILTEKE
LPEWAACTPCFRREAGAAGAGERGLVAVHQFHKVENFAFTTPNODDIAYEKMLSIVEEM
LTELKLPYRLSLLSTGDMSFTASKTIDAEVWLPGQKAFYEVSSISQCTDFQSRRSGTRYK
DSQEKLQFVHTLNGSGLATPRLLVAILENNQQADGSVVIPEVLRPYLGGLEILLPKDQ

988766 989899

CPh_0871 988766 989899
ribD-Riboflavin Deaminase
EYMEDFSEQQLFFMRRAIEIGEKGRITAPPNEWVGCVVVQENRIIGEGFHAYAGGPHAEE
LAIQNASMPISGSDVYVSLEPCSHFGSCPPCANLLIKHKVSRVFVALVDPDPKVAGQGIA
MLRQAGIQVYVGIGESEAQASLQPYLYQRTHNFFWTILKSAASVDGQVASDGWSQWITC
PEARHDVGKLRAESQAILVGSRTVLSDDFWLTARQPQGMLYPKQPLRVVLDSRGSVPPTS
KVFDKTSPTLYVTTERCPENYIKVLDSLDVPVLLTESTPSGVDLHKVYEYLAQKKILQVL
VEGGTTLHTSLLKERFVNSLVLYSGPMILGDQKRPLVGVLGNLLESASPLTLKSSQILGN
21 KVAMMETSDOVFFDTBM SLKVVWEISPOVFEPIRN

CPn_0872 989903 991216
ribAsribB-GTP Cyclohydratase & DHBP Synthase
KERIFRVACLASESVNARESMIETREEVGSANFVSLERAIEDLRAGKFVIVVDEASREDE
GDLIIAGEKITVEKMTFLLQHTTGVVCAALSQERLLSLDLEPMYKDNACRFRTPFTVSVD
AAHGYTTGVSAADRTKVVQLLADPKSKPEDFISPGHFFPLASSPGGVLKRAGHTESTVDL
MELAGLQPCGVLAELVNEDYSMMRLPQILEFARKHNIAVIPVTSIIAHRMLSDRLVSKIS
SAPLPTIYGDFTIHVVESLLEGMGHLALVKGNVAGKSNVLVRVHSECVTGDILGSKRCDC
GEQLSSAMSYIAEKGTGVLVVLRCQEGRGIGLGHKVPAYALQDNGYDTVDANLAMGFPVD
SREYGIGAQILVDLKLTTIKLITHNPQKYFGLQGFGLSITERVPLPVRISEDNEQYLRTK
QEPMSHWLDLPUCNNRVQ

CPn_9873 991188 991594
cibe_Ribityllumatine_Synthage
LSHAVTIOYNNEEYMETLKCHLSAKNLRIAIVGSCFFQAMADALVSGFQETFLKFOGGE
DJLMTIRVFGAFEIPCTIKKLLSGEKKTDAIVACGVLIQGETDHYNQIVNQVAAGIGALS
LEFCLPITLSIVAAPSAEIAWQRSGIKGRHLGVSCMTTAIEMATLFTQI

993164

CT733 hypotherical profesh LISENER FERRORDRERADMER IERTKVLVFPLALLMOCNSTOYAGDASSLOTNSOTKVK IGSEWIFOKLROYPFLLWLTESCRAPLINSTPIDMAYSEKLFNKKVPALDIAIRSMIHL

HILLIOGSROSYMOLSOILPSEEGGMTFKQFQTAHKOLLFFLN YKSFDNTLRILETAIVL RHVQCSAKAVTTFKPYFTDSCPQSFYAKALHVLRTFPELCPSYARLJFEQQEVULSLRRL GNYDSLLNLTEYPSAQLLSAWRTRRSLAILDLYLYSLDTCGDKNGSQEFYINFAPLLSML QQHATI EEAFSPYFTY RANKLGFBOTSRTDMFLVPLATLTNNLGSBEASTLAMSFKNLPSD EAENLVNSFYTVQGEHIPLTFRGLPSLVAGLSVATHGSTVSPENRLRQLYSTMLSLLVKS LRSHREMLNKQLLPQGTVLDFSETTLSSGGLDVFAESIAVRIHLNGAVSINL

993363 994022 CPn_0875

THE STATE OF THE S EM MENULYMYTER JOHN EL DE MELE THE SE VICER DEMOCTOR RESERREFAPPLOTOKUTLOVQANEKTYAEKLUEQDERDLSFVVSSAAEKSSISLALS OGEIKDALYRIREVHPLALIEALAENPALIEGMKKMQGRDWIWNLFLTQLSEVFSQAWSQ GVISEEDIAAFASTLGLDSGTVASIVQGERWPELVDIVIT

CPn_0876 994123 995517
dagA-D-Alanine/Glycine Permease
SIATGETMLYFIEQLIKLETSFCVFPMILLLGGFLTWKLRGLQFHGLKLGFNLMLQNKLD
DSSSKANEVSSYEAVAGILAGNFGTGNIAGMAVALACGGPGALVWVWLAALLGAIVQYAG
SYLGSKYRKPBGNTGEFIGGPIACLAFGMRKKILAGFFALFTIMTAFCAGNCVQVSCIVP
LCABSTPCKKLLVGILLALVVIPVLAGGNNRILRFSARVIPFIAGFYCISCGIILFQHASA
ILPAIKLICSSAFGIKAGLAGIGGYTLSQVISTGINRAVMATDCGSGMVSILQANTKSKN
PVVDGLVTLVPPVIVMVVCSITMLVLIVSGAYSSGAQGTLMVMSAFRNSLGSLGSVIVIL
AMALFGYTTILTWFACAEKSLQYMIPGRRANLWLRAITVYLIIPLGGVIDMRMIWALSDTG
FSCMVIINITALIALIALLKDVISTMEDVALIKFBCGVADDWDNIDA FSGMVILNCIALIALLKDVLSTNRDVALLKERECSVADPVRNLDA

CPn_0877 995521 995982
ybcL family
RRRIMOLLSPAFAYGAPIPKKYTCQGAGISPPLTFVDVPGAAQSLALIVEDPDVPKEIRS
DGIMIHWIVYNLSTTITNLAEGAEIFAVQGLNTSGKPVYEGPCPPDKQHRYFFTLFALDV VLPEEENVTRDQLYEAMEFHIIEQAELMGTYEKS

996660 995992

SET Domain protein
GCMSTVTTEPCSSIHISLANDWRDSQPYSLDRASELLHFRFLPSLVFSNWKVEQQIETLC
HKSEKRRLISPLAKWLGKLHKQDLLCPPAPPVSVCWINAHVGYGVFARDEIAPWTYIGEY
TGILRHRQAIWMDENDYCFRYPMPLFTLRYFTIDSGKQGNVTRFINHSEQPNAEAIGVFS EGLFHVIIRTVAPIYAGQEICYHYGPLYWKHRKKREEFIPEEE

997463 996645

YYCI-metal dependent hydrolase
YRILWKVSMQGFFPLASGSKGNSAYLGTDSCKILIDLGVSKQVVTRELLSMNIDPEDIQA
IFVTHEHSDHISGIKSFVKAYNTPIVCNLETARALCHLLDSHPEFKIFSTGSSFCFQDLE VOTFNVPHDAVDPVAFIFHYREEKLGFCTDLGWVTSWITHELYDCDYLLIESNHSPELVR QSQRPDVYKKRVLSKLGHISNQECGQLLQKIITPKLKKLYLAHLSTECNTAELALSTVSE SIASITSIAPEIALAQGITSPIYFSRLEVACPR

CPn 0880 999864 997444

CFIL U000V 797004 FISK FISK-Cell Division Protein FtsK
PMIRERKKSRHPRLPTLPLAAKASLYLFFACFSGLSLWSFHRDOPCTONWIGLLGWSFSS
FLLYFFGAAAFIPLYFLWLSFLYFRRTPRPLFFYKAAAFLSLPFCSAILLSMLSPVGTL
PALLDTRLPKFILGNNPPVSYVGGIPFYLFYEGGSFCLKHLIGSVGTALIFGFVMLFSVL YLCGGIALLKKKTFQDGVKKAFCSFFQTCFKNLKKLINRRNYLPKPSVPFVSKNPFSCTK SQPSPRRVSETIILDGSISPLPQEEIPGSKKESFFLTPHPCKRFLTKFVEPQENKAKEGK SOPSPRRVSETI ILDGSISPLPQEEIPGSKKESFFLTPHPCKRFLTKFVEPQENKAKEGK
TIALSSTPTVVRESKGKERAALPKLKSLAVPENDLPQYHLLSKNREARPESLQAELERKA
LILKGTLTSFGIDADLGNICSGPTLAAFEVLPHSGVKVQKIKSLENDIALKLQASSIRII
APIFGKAAVGIEIPTPFPQAVNFRDLLEDYOKTNRKLQIPLLLGKKANGDNLWADLATMP
HLIIAGTTGSGKSVCINTIVMSMIMTTLPSEIKLVIIDPKKVELTGYSQLPHMLSPVITE
SREVYNALVWLVKEMESRYEILBYLGLRNIQAFNSRTRNKTIEASYDREIRETMPFWGI
IDELSDLLLSSSQDIETPIIRLAQMARAVGIHLILATQRPSREVITGLIKANFPSRISFK
VSNKVNSQIIIDEPGARLMKNGGMUVLLPSVFGTTRAQGAYICDEDINKVIQDLCSRFP
TQYVIPSFHAFDDSDSDNSGEKDPLFAQAKTLILQTGNASTTFLQRKLKIGYARAASLID
QLEEARIIGPSEGAKPRQILIQNPLEG

CPn_0881 1005646 1006209 No robust homolog present in Genebank/EMBL as of 11/7/98 NKKFAVHMPVPIDNSSRNLQEVPESLEDLEQHAEESPTHQSAESSSLQLSLASSAISSRV EQLSSLVLGMENSDFSSLRDVPIFSAIYESSTHTEVVTPLVGVGYINGSQGGYYDTQRES LHLSQLLGSRRVEVVYNQGNFMEASLLNLCPRRPRRDPSPISLALLELWEAFFLEHPPGS TFNPIFFW

CPn_0882 1006169 1007404 No robust homolog present in Genebank/EMBL as of 11/7/98 NTPOVALLIQYFFGNOAFYVREALRITHAQNIVLVGICPSLYPEHPRSFYYRVSGDIGS RFDDRGFVNSGVETLPYSSGSFGIFWISFTDPTFNFAIWNTFMRTAGINEVSRPMTQDTE RFDURGYWNSVETEPYSSGRGIFWISTDETTRIATWTFMRTAGINEVSREMTQDTE TSLIEMROLSEQGEANNTOSLEGGESLMGIVGHTVGGVSHTVTSSPNIFYRIGTLLGLPE TLAEAEENPTFFNSTIDSLAEIMMNLVRISDAVSIFWIFFIVDTTYMGVLLAVCIGFFGI NGICSTFLMLTNPRSRRDRWRNLRIMVLCYRSLGSGMNLFDLSNNVRMAARRHVTSCTVA LYAMVTLFGWTVAIQDALQYGFFSVRDAFYRYCLRHRYCLTQRNEDSLQTTGTRFQVTRT HLEDQQMVASILNLSVFGLFFGFVGLMTTFGGLEISPSCRWDAANNRTVGIF

1008904 1007573

CPH_0883 1008904 1007573
dmpP/nqtb-Phenolhydrolase/NADH ublqulnone oxidoreductase
LYELFIKSGIFIYMTWLGGLYFIGIASLIFCAIGVILAGVILLGRKLFIKVHPCKLKIND
NEELTKTVESGGTLLVGLLSSGIPIPSPCGGKATCKQCKVRVVKNADEPLETDRSTFSKR
QLEEGWRLSCQCKVQHDMSLEIEERYLINASSWEGTVISNDNVATFIKELVVAVDPNKPIP
FKPGGYLQITVPSYKTNSSDWKQTMAPEYYSDWEHFHLFDQVIDNSQLPADSANKAYSLA
SYPAELPTIKFNIRIATPPFINGKPNCEIPWGVCSSYVFSLKPGDKITVSGEYGESFMKD
DDRPLIFLIGGAGSSFGGRSHLDLLLHVHSKREIDLWYGARSLKENIYQEEYENLERQFF NEHYHLVLSEPLPEDIAAGWDKDDPTYTNFLFRAFNLGQLGRLDNPEDYLYYYCGPPLHN COLLKILODYGVERSSTILDDFGS

CPn_0884 1009368 1009009
CT741 hypothetical protoin
CDX MEGRIVTCELECEREPEABEEAAGGRNTEVOPAVMEATATEEPYETEWRPEQKRR
KAMEKPKNDLAKGPAVTAMGTIGTVDDTPEREVILHTAGGRVEVEVEKGATGETERPNDNKG

1011288 1010908 hctA-Histone-Like Developmental Protein

RTLEMALKOTAKKMKOLLOS IOHOLAKAEKONKAAAORVRTOS IKLEKVAKLYRKESIKA EKAMBURKE TEALA OTEMA PERANAKKONKAAAORVRTOS IKLEKVAKLYRKESIKA

CPI_0887 1011692 1014157
CHLTR possible phosphoprotein
MKKLYHPTLFLRPLIRLSLIFALSLTLISGNFPQKSFGHCCADMHSALISGKNCEELFA
DFIERVLADRETLTARDWGTVVVLVREYLLKCIRKGDCDYGVKILQKLLALRLPKDARKD
COLUMHRINPEQAPLEDVVDQLFTIGCHESLOPHLLFELYTVTLHSGYENRKQDMLLAKE
QGDYKRAIELAKELVAALEKGSCSHPEIVQIEKTFLQKTLLALQIKVAQEAQESCDALL
TPYCLSEIAYTEAMDALVLRIARGEVSRTNEVDSVLLSHALQHLPFAREKAIPELEVLID
HGAYLESTLLYYAYFSLLELYHQNKDFASLERLLEKGDAVFVPEHPYFPEYGFFLGAYFY
AKGKYESAEKVFLQIIDPAVKLGATFARAYEYLGSLIAYVQNHYEKAEEYFLRAYKSWGRE
ESGIGLFLAYAVOKKKTACEDMLYHPKFSFTYRHLLDSLGLSVFHGENKGSSAIGNVHR
AVPELSEIYSRCIVDMIKYRNYTYTHPIIELAYNQVRNLEKRNLEEICRDAQDPEYDKAL
AFWGALQSGASVPRSLIESSDVDEARITIRCYEALYFHNPDAIAMLPQAFSEECNSWGTA
LRLWMTLVRPKGAPNHAKYWDHLVLRPHGDSLYFFGYDLQEYLIGKEDALKHLSVFAELF
PKSSLLSLVYYLQGYSESSALRKVGWFVVAALEEFTEISWSGEHMKTWAYIYYMVKLDLAD
TYISLGNFSQAVHILEEVKEEMQVASHPKLHFLKGEDCYLAMELRWVEGLAYAYFQLHET
AHLSNHLLEHVEKNLISPRSYRDYYGESLQRTLGLCQRFLGV

1014119 1015441

CPT_U888 1015441 1014119
hemG-protoporphyrinogen Oxidase
AERFCVKRAIIIGAGISGLAAGWALHKKFPQAEILVLDKEAYAGGFVRTESPQGFSFDL
GPKGFLTRGDEFYTLKLIHELGLQNSLIFSDRAAKNRFVYYRGKAHKISTWTLLRKGLLP
SLIKDFRAPCYTQDSSVQDFLKRHSSQNFTSYILDPLITAIRAGHSSILSTHMAFPELAK
REASSGSLLRSYLKNRSPKKSKTDRYLASLSPSWGTLITTIQEKLPATWKFSTSVTHIDC REASSGELIS LINNSTANTINI LEGISTS OF THE REAL PROPERTY OF THE PR

CPn_0889 1016841 1015462
hemN-Coproporphyrinogen III Oxidase
FEMENVNFKFLEGLHOPAPRYTSYPTALEWEPSDAAPALLAFQRIRENPQPLSLYFHIPF
CQSMCLYCGCSVVLNRREDIVEAYINTLIQEMKLVVETIGFRPQVSRIHFGGGTPSRLSR
EUFFLLFDHIHKLFDLSHAEEIAIEVDPRSLRNDMEKADFFQNVGFNRVSLGVQDTQADV
QEAVRRRQSHEESLKAYEKFKELAFQSINIDLIYGLPKQTKESFSKTIQDILAMYPDRLA LFSFASVEWIKPHOKAMKASDMPSMEEKFAIYSOSRHLLITKAGYOAIGMDHFSLPHDPLT LAEKNKTLIRNFOGYSLPPEEDLLGLGMTSTSFIRGIYLONAKTLEEYHMTVLRGTFATV KSKILTEDDRIKKWALHKLMCTFIINKEEFFNLFGYEFDTYFIESRDRLISMETTGLIHN SEJSLKVTPLGELFVRVIATAFDHYFLNKVSKKECFSASI

CH-0890 1017829 1016819
henE-Uroporphyrinogen Decarboxylase
stilhwdsmsaffdlksotashppiwllrovdrymppyQelkgsoslktfphnteaive
atticpsllhydaailfadilsildgfavtydfapgprigfspeopffffsdpgtiffyl
LDAGRTLKOKLEVPLIVFAASPFTLACYLIDGGASKDFSKTMSFLYVYEKFDQLISTII
EGTAIYLKTOMDAGAAAVQLFESSSLRLPSALFTRYVTEPNRRLIAKLKEQAIPVSLFCR CFEENFYTLOATOADTLHPDYHVDLHRIOKNLMLSLOGNLDPAIFLLPQEKLLHYVEAFL VELRTYPNFIFNSGHGILPETPLENVOLVVSYVQROL

MEG-Transcription-Repair Coupling
MEG-Transcription-Repair Coupling
MEG-Transcription-Repair Coupling
MEMMOFNEVNLDFSISKEFKEETLPLLLENIHGATAFLAAKMFHDCRASVIMITTPAR
MEDDLFENLRTFLDQAPVEFPSSEIDLSPKLVNIDAVGKRDHLLYSLNQHRAPIFCVTTLK
ALÆKTRSFQATSQQHLDLAVGDVLDPEATTELCKSLGYSCVMLTSEKGEFSCRGGIVDI
FPLSSPEFFRIEFWGEKIISIRSYMPSDQLSTGKVSKISISPAYTEEASGGNYSHSLLDY
FSTÆPLYLFDNLEILEDDFADISCTLSSLPDRFFSICTLYRISTSNQVYFSETPFPNVK
NLKENRVIIEAFHRNMEASRQAIPILYPEQIIONDENPLLAFLQHLØEYMPPHGKPLKLA
IYSEKTKSLKEARALAETVARGDVEIYEKTGNLTSSFALVNEAFAAISLSEFASTKVLRR
VORORTHFSVTTEEVFUPIPGGTVVHIHNGIGKFLGIEKKNHHLNETDYLVLEYADKARL
YVPSNQAYLISRYVGTSDKAADLHHLNSSKWKRSRDLTEKSLIVYAEKLLQLEAQRSTTP
AFVYPPHGESVIKFAETFPVEETFDQLKTIODIYNDMSPKLMDRLICGDAGFGKTEVIM
RAAVKAVCDGHRQVIVMVPTTILATOHYETFKERMAGLPIEIAVLSFSOAKVOKLICEO AFVYPPHGESVIKFAETFPYEETPDOLKTIDQIYNDMMSPKLMDRLICGDAGFGKTEVIM
RAAVKAVCOCHRQVIVMPTTILATOHYETFKERMAGLPIEIAVLSRFSOAKVOKLICEQ
VASGOIDIIICTHKLINKSLEFKNPGLLIIDEEQRFGVKVKDNLKERYPMIDCLTVSATP
IPRTLHMSLSGARDLSVIAMPPLDRLPVSTFVMEHNTETLTAALBHELLRGGGAYVIHNR
IESIYTLAETIRNLIPEARIGVAHGGMGAEDLSNIFTKFKNQKTDILVATALIENSIDIP
NANTILIDHADKFGMADLYQMKGRVGRWNKKAYCYFLVPHLDRLSGPAAKRLAALNKQEY
GGGMKIALHDLEIRGAGNILGTDQSGHIGTIGFNLYCKLLKKAVSALKKHTSPLLFNDDV
KIEFPYNSRIPDTYIETGSMRIEFYQKIGNAESSEELTAIQEEMBDRFGPLPQEICFULFA
LAEIRLFALQHGISSIKGTANALYVQKCLSKSEQTKKTLPYALSPTPELLVKEVIESIER
GFLINAS

1023673 1021046

alaS-Alanyl tRNA Synthetase

Alas-Alanyl trna Synthetase
EFFFMLSNTIRSNFLKFYANRHHTILPSSPVFPHNDPSILFTNAGMNOFKDIFLNKEKVS
YSRATTSQKCIRAOGKHNDLDNVGHTSRHLTFFEMLGNFSFGDYFKAEAIAFAWEVSLSV
PNFNEGIYATVHEKDDEAFALWEAYLPTDRIFRLTDKDNFWSMANTOPGGYCSELLFDR
GPSFGNASSPLDDTDGERFLEYWNLVFMEFNRTSEGSLLALPNKHVDTGAGLERLVSLIA
GTHTVFEADVLRELIAKTEQLSGKVYHPDDSGAAFRVIADHVRSLSFAIADGLLPGNTER
GYVLPKILRRSVNYGRRLGFRNPFLAEIVPSLADAMGEAYFELKNSLOQIGKYLTLEEES
FFKTLDRCONLLOQVLKSSSSSSCIGGEDAFKLKDTYGMPIDEISLLAKDYDYSVDMDTF
HKLEGDAKERSHKNVVQSQCTSESIYNFELHTTSEFIGYDHLSCDTPIEAILCYDHIVSSL
OEKOEGAIVLKVCPFYAEKCQVGLOGEIFGSEGTFIVTHTTSRKAGLIVHKCRISQGSL
TVEAAVTAQVNRYRRKEIANNHTACHLLHKALEITLGDHIRQASSYVDDTKIPLDFTHPO
ALSPEDLALTETLNRSGIRENEPPVLIPEALYSCHMMSSGIKQFTORKYDDVFTVSAGHS
HELCCTTHAEATVDIGFFRITKFHAVAMSIRRIEAVTGCKAEATVHQOGEVLEEIATLLQ
VPHDJIVSHITATLOEPKQQKKLHELENSLIQFKGKAETVHQOGEVLEEIATLLQ
VPHDJIVSHITATLOEPKQQKKLHELENSLIQFKGKAETVHQOGEVLEEIATLLQ
VPHDJIVSHITATLOEPKQQKKLHELENSLIQFKGKAETVHQOGEVLEEIATLLQ
VPHDJIVSHITATLOEPKQQKKLHELENSLIQFKGKAETVHQOGEVLEEIATLLQ
VPHDJIVSHITATLOEPKQGKATHTEKNIKYILSTROEPKGAGAGTATAQDLLPAVLTPCG
GRWYSKDQGAQGGAPALPATLVLNETLWWHITDLII GRWGGKDQBAQGBAPALPATEVLNETEWQWIGTQLI

CPn_0893 1023822 1025898

t kt B Transket olase

EFLAFCLGIGYSCCFYIEGLOGLIMINKELDIGILGKIAGAIKQIGIESIOKASSCHPGL
PLOCAELAAYLYGYVLRQNPPDFHWINRERFYLJAGHGSALLYGCHLAGFDVSLEDLQE
FROLHSRTPGHEFYGETVGVEATTOPLAGATICRAVGHALLGMMLESRFNRROHEIFNGKI
YCLAGDCCFMEGVSHEVCSFAGSLNLNNLVVIYDYNNVVLDCYLNEISVEDTKKRFEAYG
WDVYEIGCYDFTHIHETFSSIKROQERFYLVIAHTIIGHGSPKEGTNKAHGSPLGVEGTH
ETKQFWHLPEEKFFVPPAVKNFFAHKICGDRRAQEOMLDEVRVWSKQFPELHEEFVALTS
HKLPKNLESLVQSVEMPDSIAGRAAGNKLIQVLVQHIPYLIGGSADLGSSDGTWIANEKV
IHTYDESGRNIKYGVREFGMATIMNGLAYSOVFRPFGGTFLVFSDVMRNAIRLAALSKLP
TOFTHIL TOFTHI OF THE TOTTHIN TO THE TOTTHIC TOTTHIN TO THE TOTTHIC TOTTHIN TO THE TOTTHIC TOTTHIN TO THE TOTTHIN TO T

1026823 1025888

cei_gos4
amn-AMP Nucleosidase
PRNDKNAKNLRKHYKGERVSKHTSESRIAQDMLERYSGSSVKQFCPYLLLTNFSYYIQT
FAKLHGVPVFEGSMFSAAHAPHLKTSILDFKLGSPGAALTIDLCSFLPDLKAALMLGMCG CLRSHYQVGDYFVPVASIRGEGTSDAYFPPEVPALANFVVQKATTEVLEDKKANYHIGIT HTTMIRFWEFNKKFRKKLYETKAQSAEMECATLFAAGYRRNLPIGALLLISDLPLRKEGI KTKSSGNFIFNTYTEDHILTGQEVIENLEKVMLKRAASDHKKDQQYRGLPHMEVGEADDT MASGSETSDSDY

1026973 1027557

efp-Elongation Factor P EIDCFMVRVSTSEFRVGHIEIDCOPYLILQNDFVKPGKGOAFNRIKVKNFLTGRVIERT VKSGESVETADIVERSMRLLYTDQEGATFMDDETFEQEVVFWEKLENIRQWLLEDTIYTL VLYNGDVVAVEPPIFMELSIAETAPGVRGDTASGRVLKPAVTNTGAKIMVPIFIDEGELV KVDTRTGSYESRVSK

1027574 1027822 CPn_0896

CTT53 hypothetical protein
EKYFFFTVRNMEAKKIKELSKEAQLLKKLREKSRVLDEKNKRKAWVAKLVAMPESIREIE KEERVETPQLFQAIAEKILEEGV

1028794 1027853

(phosphohydrolase) (phosphonyarolase)
NFSLDSNTYDQKNKSNPRPMQEKPRHVHRIIHISDVHFHVLPVNPVHCFNKRLKGLLRKV
FGLVHFQATTIGQRFPKVVRSLGADSVCITGDFSLTAMDGEFLLAKHFVETLAKHSSVYL
LPGNHDVYTLKSLAQQTFYTHFPNDQLQQNKVSFHKITDHWWLILLDCSCLKGWFSANGV
VHLAQISAIETFLLSLSPEENVIIANHYPLLSSQNPSHDLINNTHLGNVLKKYPKVRLYL
HGHEHQAAVYNCADTSPSYILNSGSISLPTNSRFHVIDLYPEKYQVHTMILKNLLDFDAP

CPn_0898 1030511 1028904
Mitochondrial HSP60 Chaperonin Homolog
TKKRLGSVKILRLLGVCMSEQEKLSNYNADKKLFSGIDKLFQIVKGSYGPKQSLSPTSFF
KERGFYAISQTELSNSYENLGVOFAKAMVNKIHKEHSDGATTGLILLHAILQESYAALEK KERGFYAISGTELSNSYENLGVDFAKAWNKIHKEHSDGATTGLILLHAILQESYAALEK GISTHKLIASLKLQGEKLQEALQQSWPIKDALKVRNIIFSSLHMPTIADHFYNAFSVVG PEGLISITKERENDKTSMDVFQGFKIPAGYASTYFVSDTASRLTRIAHPLILITDRKISM IHSLLPLLQEISEQNOHLIIFCEDIDPDVLATLVVNKLQGLLQVTVVTIPQLSTTNQELA EDIALFTCFHICPCQEASHVLAPEMVTLGSCLSIEISESGTTLIGGLHIPEVLTLKTRQL AEEIRTTSCLETKKRIIKSTNRLQSSVAILPTDEDNEPLYTLALKIMESALSRGYVFQGG VALFYASLTLGTPKDDADENSIAISLLQKACCAPLKLLATNADLDGDAVIAKLSSLGTTS LGISVFSREIEDLIAGGILDSLATTSTILAQALDTAILVLSSKILILENQYEISTL

1030848 1032215 CPn_0899

CPn_0899 1030848 1032215

muff-Muramoyl-DAP Ligase

NHRCCRQNYMRAMLLEDWVSLMLSDVSCPKCDKKITGFAIDSQQVQPGDLFFALPGNATD

GHQFLKHAATAGAVAAVVSHDYQGDSFGLELIRVDDTKSALQEAGSNQCNLFQGTLVGIT

GSVGKTTTKEFSKTILSSIYKTHASPKSYNSQLTVPLSLLMAEGDEDWMILEWGVSEPGN

MQDLLRIVQPEIAVITHINDQHAMHFPQGIQEILKEKSY ILQKSKLQLLPKDSPYYLDLR

SCSPTAEKFSFSFNDPLADFCYKAISGDSVVIQTPEENYCLPIAFSYKPAYTNLLIAVAL

SWILEVPEGGVIRSLEPELKLPPMRFEHSWRNGMQVINDAYNACPEAMIAALDALPLPSDG

GKIILILGHMAELGRYSEEGHALVAEKAASRGDMIFFIGEKWIPVQSVLKSYSCEVSFFS SAQDVKDILKQVARYGDVILLKGSRALALESLLACF

CPn_0900 1032208 1033281
mraY-Muramoyl-Pentapeptide Transferase
LVFNFLGASMIPLIPMFLKQSLFFSLALTGMTTLVLTVALGVPVMKWLKRKNYRDYIHKE
YCEKLEMLHKDKAEVPTGGGVLLFISLIASLLVMLPWGKFSTWFFIILLTCYAGLGWYDD
RIKIKRKQGHGLKAKHKFMVQIAIAAFTLIALPYIYGSTEPLWTLKIPPMEGMLSLPFWL
GKVFCLGLALVAIIGTSNAVNLTGGLOGLAAGTMSFAALGFIFVALRSSTIPIAQDVAYV
LAALVGACIGFLWYNGFPAQLFMGDTGSLLLGGLLGSCAVMLRAECILVVIGGVFVAEAG SVILQVLSCRLRKKRLFLCSPLHHHYEYQGLPETKIVMRFWIFSFVCAGLGIAAVLWF

CPn_0901 1033239 1034537
murD-Muramoylalanine-Glutamate Ligase
FCMRRSRYSGCIMEIDMCORILILGTGTTGKSVARFLYQQCHYLIGADNSLESLISVDHL
HDRILMGASEFFENIDLVTRSPGTKPYHPWEQAVSLKIPVVTDIQVALKTPEFQRYPSF
GITGSNGKTTTTLFLTHLLNTLGIPAIAMGNIGLPILDHMQQPGVRVVEISSFQLATQEE HIPALSGSVFLMFSRNHLDYHRNLDAYFDAKLRIQKCLRQKTFWWEECSLCMSYQIYS
EEIEEILDKGDALKPTYLHDDRDNYCAAYALANEVGWVSPEOFLKAIRTFEKPAHRLEYLG
KKDCVHYNDSKATTVTAVEKALMAVGKDVIVILCOKDKGODFPALASVLSQTTKHVIAM
GECROTIADALSEKIPLTLSKDLQEAVSIAQTIAQEODTVLLSPGCASFDQFQSFKERGA YFKLLIREMOAVR

CPn_0902 1034507 1035241
nlpD-Mutamidase (invasin repeat tamily)
AVDQPNAdSEVNMNRRDMVITAVVVNAILLVALFVTOKRIGVKDYDEGFPNFASSKVTQA
VVJEKVIEKPVVARVPGRFIAKETLAAQFIESKPVIVTTPPVPVVVGETFEVPTVAVPPQ
EVRETVKEGAPYATVVVKKGDFLERIAPANHTTVAKLMQINDLITTQLKIGQVIKVPTS ODV:NEKTFOTOTANPFNYY IVQEGDSFWT IALRNH I PLIDDLLYMNDLDEYKARRLKPGD

OLPTP

CPH_0703 10 0.745 1036417

ETHW COLL DIVIL KON PROFESS IN THEW

ACLLYMMEWROPER CHURTHWORTPOSSERVLORGERGOTHRAL (POVTYL (LACOV ACLLYMMEWROPER CHURTHWORTPOSSERVLORGERGOTHRAL (POVTYL (LACOV ACLLYMMEWROPER CHURTHWORTPOSSERVLORGERGOTHRAL (POVTYL (LACOX) AVERTHWORTPOSSERVLORGERGOTHRAL (POVTYL (LACOX) AVERTHWORTPOSSERVLORGERGOTHRAL (POTTYL (LACOX) AVERTHWORTPOSSERVLORGERGOTHRAL (POTTYL (LACOX) AVERTHWORTPOSSERVLORGERGOTHRAL (POTTYL ACC)

ECHOPOGROWOLFILA VILLORGER AVERTHYR (LACOX) AVERTHWORTPOSSERVLORGERGOTHRAL (LACOX)

PETTYROLOGICAL COLORGERGOTHRAL (LACOX)

ELORGERGOTHRAL (LACOX) AVERTHWORTPOSSERVLORGERGOTHRAL (LACOX)

ELORGERGOTHRAL (LACOX) AVERTHWORTPOSSERVLORGERGOTHRAL (LACOX)

ELORGERGOTHRAL (LACOX) AVERTHWORTPOSSERVLORGERGOTHRAL (LACOX)

ELORGERGOTHRAL (LACOX) AVERTHWORTPOSSERVLORGERGOTHRAL (LACOX)

ACLLYMMEWROPER (LACOX) AVERTHWORTPOSSERVLORGERGOTHRAL (LACOX)

ACLLYMMEWROPER (LACOX) AVERTHWORTPOSSERVLORGERGOTHRAL (LACOX)

ACLLYMMEWROPER (LACOX) AVERTHWORTPOSSERVLORGERGOTHRAL (LACOX)

ACLLYMMEWROPER (LACOX) AVERTHWORTPOSSERVLORGERGOTHRAL (LACOX)

ACLLYMMEWROPER (LACOX) AVERTHWORTPOSSERVLORGERGOTHRAL (LACOX)

ACLLYMMEWROPER (LACOX) AVERTHWORTPOSSERVLORGERGOTHRAL (LACOX)

ACLLYMMEWROPER (LACOX) AVERTHWORTPOSSERVLORGERGOTHRAL (LACOX)

ACLLYMMEWROPER (LACOX) AVERTHWORTPOSSERVLORGERGOTHRAL (LACOX)

ACLLYMMEWROPER (LACOX) AVERTHWORTPOSSERVLORGERGOTHRAL (LACOX)

ACLLYMMEWROPER (LACOX) AVERTHWORTPOSSERVLORGERGOTHRAL (LACOX)

ACLLYMMEWROPER (LACOX) AVERTHWORTPOSSERVLORGERGOTHRAL (LACOX)

ACLLYMMEWROPER (LACOX) AVERTHWORTPOSSERVLORGERGOTHRAL (LACOX)

ACLLYME (LACOX) AVERTHWORTPOSSERVLORGERGOTHRAL (LACOX)

ACLLYME (LACOX) AVERTHWORTPOSSERVLORGERGOTHRAL (LACOX)

ACLLYME (LACOX) AVERTHWORTPOSSERVLORGERGOTHRAL (LACOX)

ACLLYME (LACOX) AVERTHWORTPOSSERVLORGERGOTHRAL (LACOX)

ACLLYME (LACOX) AVERTHWORTPOSSERVLORGERGOTHRAL (LACOX)

ACLLYME (LACOX) AVERTHWORTPOSSERVLORGERGOTHRAL (LACOX)

ACLLYME (LACOX) AVERTHWORTPOSSERVLORGERGOTHRAL (LACOX)

ACLLYME (LACOX) AVERTHWORTPOSSERVLORGERGOTHRAL (LACOX)

ACLLYME (LACOX) AVERTHWOR

OLEGAALAMVITUIISMQAFMNLGVVSGLLPSKGVNLPFFSQGGSSLIANMCGVTLLLKV YDEENSKSGU/CRRFRRPHCPSSLGKGSFFS

1036320 1037396

CPn_0904 1015320 1917336 murG-peptidoglycan Transferase RYMMKKIRKVALAVCGSCGHIVPALSVKEAFSREGIDVLLLGKGLKNHPSLQQGISYREI PSGLPTVLNPIKIMGRTLSLCSGYLKARKELKIFDPDLVIGFGSYHSLPVLLAGLSHKIP LFLHEONLVTGKVNOLFGRYARGIGVNFSPVTKHFPCPAEEVFLPKRSFSLGSPMMKRCT METTHEVVALES WOLLDE VESSALVELDE OF THE TOTAL SESSANDELT CONTROL OF THE TOTAL SE

CPn_0905 1037400 1039835
murC&ddla-Muramate-Ala Ligase & D-Ala-D-Alam Ligase
WHYMKGTPQYHFIGIGGIGMSALAHILLDRGYEVSGSDLYESYTIESLKAKGARCFSGHD
SSHVPHDAVVVYSSIAPDNVEYLTAIQRSSRLLHRAELLSQLMEGYESILVSGSHGKTG
TSSLIRAIFQEAQKDESYAIGGLAANCLMSYSGSSKIFVAEADESDGSLKHYTPRAVVIT
NIDNEHLNNYAGNLDNLVQVIQDFSRKVTDLNKVFYMGDCPILKGNVQGISYGYSPEQL
HIVSYNQKAWQSHFSTTFLQGEYDDIELNLPGQHNAANAAAAGGVALTFGIDINIIRKAL
KKFSGVHRRIERKNISESFLFLEDYAHHPVEVAHTLRSVRDAVGLRRVTAIFQPHRFSRL
EECLQTFPKAFQEADEVILTDVYSAGESPRESIILSDLAEQIRKSSYVHCCYVPHGDIVD
YLRNYIRIHDVCVSLGAGNIYTIGEALKDFNPKKLSIGLVCGGKSCEHDISLLSQHVSK
VISDPFYNDYSYFIINGGIMBTGGDPPHLIEFTGGDSPLSSEIASGLAKVDCLFPULHGP YLSPETYDVSYFIINRGGLWRTGKDFPHLIEETYGDSPLSSEIASALAKVDCLFPVLHGP
FGEDGTIQGFFEILGKPYAGPSLSLAATAMDKLLTKRIASAVGVPVVPYQPLALCFWKRN
PELCIQKLIETFSFPMIVKTAHLGSSIGIFLVRDKEELQEKISEAFLYDTDVFVEESRLG
SREIEVSCIGHSSSWYCMAGPNERGGASGFIDYQEKYGFDGIDCAKISFDLQLSQESLDC
VRELAERVYRAWGKGSARIDFFLDEEGNYWLSEVNPIPGMTAASPFLQAFVHAGWTQEQ

IVDHFIIDALHKFDKQQTIEQAFTKEQDLVKR

CPn_0906 1040514 1039915 CT763 hypothetical protein kWGSEVLELVNDSQLSREASAFRLDIDFFILNIYPFFRNFKNIELCFFLSISQFNLDFME EFVAYIVKNLVTNPEAVEIRSIEDEDNESIKLEIRVAAEDIGKIIGRRGMTIHALRTILR RVCSRLKKKVQIDLVQPENGTDVIADQDYICDNDSSNSTEDTFGESDTCCSGHCHYDEDL NQEEQEEGNMHHSCECSNHH

CPn_0907 1040816 1040445
*cutA Periplasmic Divalent Cation Tolerance Protein CutA (C-Type Cytochrome Biogenesis Protein)
FAFSKFLIIKSSMTAVLILTSFPSEESARSLARHLITERLASCVHVFPKGTSTYLWEGKL

CESEEHHIQIKSIDIRFSEICLAIQEFSGYEVPEVLLFPIENGDPRYLNWLTILSYPEKP

CPC_0908 1041607 1040780
CT754 hypothetical protein
ILATLIFMIIIKINELMIREFFKTLFPFGPQYSLCYASILIVLSSLVCVPTFCWLFLPELS
LSKTNPSPIRNLFLVSSTLSKVPPTAIAEHLRLSADAPTYLHEFSIKEAESSLHALGIFS
SLVIEKSPDNKGITIFYTLOTPIAVGNRSNTLCNLEGSCFLGQPYFPSLNLFQIFFSQE
DLKWGKLPFEKMLFTKILLKELAMESPKIIDLSLSDAYPGEIIVTLSSGSLLRLFIKTLD RALDLYKHMKKSPVIESEKQYVYDLRFPNFLLLKAL

CPA-0909 1041592 1041966
rsby-Sigma Factor Regulator
IISUFFTRFLLERLLMMLSAKEYGDIIVIYLQGSLDAVSVPSVQEYLEQFIQKKHLKIAL
NFTDVSYISSAGIRLLLSNFKLVQSLGGKMCLCCVKESVTEVMRIAGLDQLILLCQSEQE
CLSKL

1041970 1043004

CPn_0910 1043004
miaartrna Pyrophosphate Transferase
FLYNLPFEFEFNTTSSPECDVCLDPOKLFVKLFKRTIVLLSGPTGSGKTDVSLALAPMID
GEFVSVDSMQVVGSMDIGTAKVSLKARGEIPHHLIDIRHVQEPFNVVDFYYEAIQACQNI
LSRNKVPILVGGSGFYFHAFLSGPPKGPAADPQIREQLEAIAEEHGVSALYEDLLLKDPE
YAOTTKNDKNKIIRGLEIIQLTCKKVSDHEWDIVPKASREYCCRAWFLSPETEFLKNNI
QMRCFAMLQEGLLEEVRGLLNQGIRENPSAFKAIGYREWIEFLDNGEKLEEYEETKRKFV
SNSWFYTKKQKTWFKRYSIFRELPTLGLSSDAIAQKIAKDYLLYS

CPn _______911 1044079 1042985

CPT=#911 1044079 1042985
Fe=#91cluster oxidoreductase
SLEEAIFNVNYFMNLCKRISFEEGLELFVSSPIERLQERADAIRKERYPSNEVTYVLDAN
PNYTNICKIDCTFCAFYRKPKSPDAYLLSFDEVRSLLQRYVSSGVKTVLLQGGVHPGLGI
DYLEELVRITVQEFPSIHPHFFSAVEIEHACRVSGISIEQGLQRLWDAGQRTIPGGGAEI
LSERVRKIISPKKMPGGWINLHKLAHLMGFRTTATMMFGHVENPEDILIHLQTLRDAQD
SCPGFYSFIPWSYKPGNTALRRNVPQQASIETYYRILALGRIFLDNFDHVAASWFGEGKS
LGAKALHYGADDFGGVILDESVHKATGWSIQSSEEEICNIIRSEGFIPVERNTFYQHISC

CPn_0912 1044120 1045760
CT768 hypothetical protein
VVIMDNSDNSFHTLETEQGSFLNDELAVEEVASTESTEISDATLCFAEKKVAFILNKMRE
ALTGSSQGSDLRLFWDLRKQCLPLENEIEDTAKRADHWRCYIELTKEGRHLKGLQDEEGS
FVVQGIDLAITCLEKDILKFOGETEDKIFKDREDNFLESQALDKHQAFYKQHHTSLLWLS
SFSSKIIDLRKELINVGMRMRLKSKFFQRLSNLJNQVFPKRKELIEKVSQTFAEDVDAFV
AKYFIGSDKETLKKTVFFLRKEIKHLQHAAKRLFVSSHVFAETRLKLSKCWDQLKGMEKE
IRQEOQRLRVVSAENSKEVRGMLAEVSSLLIECNDLSKVRKDLEGISKKIRALDLTHDDV
ISLKKEMQQLFDQLREKQDAAEHSYQEQLAKDKQVKYEAARSLAERITTFSKTCSEGNIT
SESREEWQTLKELLGKMSFLPPPEKISLDNQLNLALQTIVNFFEEQLLSSPDSREKLVNM
RQVLKQRRERRQELKDKLEQDKKLLGSSGLDFDRAMQYSALVEEDKRALEELDASILELK
QOIQQLL QQTQQLL

CPn_0913 1045709 1045945
No robust homolog present in Genebank/EMBL as of 11/7/98
RICKKYPRIEATDGAIAMRRNCIYAFDLDGTLLKGNOOWGFYCYGLLQGLFGYKTLPPYI YRPFREKEFFCIFHPSIIR

1045909 104639 Cltn 0914

NO TORIST HORROTOR PRESENT IN GENERAL/EMBL AS OF 11/7/98
VEFWDLESEYYSTVTRILASVECDDLYEVALINEVETLES DEPARVLEKLEEAFADTESO
VILESSIDE FURIE ALOQUATISAVASCYPLOSAFOT LYKKCLTSOKKAQ (LGYTKKINO
ARSHTESOBILLDLEFEMLGEEKTVVRPOGREKKMAYEYYWNIV

Oni_0915 1046401 104681*

ybeB-toldp_supertamily_ortholog FHLKKSSTLSWITEYPKAGFMDSFCFDLLKVAAKAIDDKKGNNLVVLDVRTISEFTDYFV FVBCSVNVHVKALARTIVEELKKQKVSPLHVEGITDGNWVVIDYGFIVVHVFVSEIRGKY RLEELWKDGFIVTSKLLAS

1046813 1048084

CPH_0916 1046813 1048084
tabF-acyl Carrier Protein Synthase
LLNCVRVYMSKKRVVVTGFTVVSCLGNEVDTFYDNLLAGVSGVRPITSFPCEDYATRFAG
AN ELLET CYMST. AND THE TOTAL AND PRODUCTION OF THE TOTAL TOTAL COMMANDERS OF THE TOTAL COMMANDERS OF

CPn_0917 1048054 1048539 hydrolase/phosphatase homolog FNDIILEVCTLVMMKTKYEVSFGVIPIKFFGTPDKNTLKACFICHTRGKHWGFPKGHSED KEGPQEAABRELVEETGLSVVNFFPKVLIEQVSFNNEEQVFVRKEVTYFLAEVRGDIHAD PMEICDSQWLSLQEGLRLLSFPELRDLTVEADKFINNYLFSS

1049232 1048579

ppa-Inorganic Pyrophosphatase
ELLMSKREYLVANPMHSPTLTQDNYESLCCYIEITPYDSVKFELDKATGLLKVDRPQKFS
MFCPCLYGLLPQTYCCTASGNYSGGTRREGIGGDKDPLDVCVLTEKNIHHGNILLQARP
IGGLRIIDSGEADDKIIAVLEDDLVFAEIEDISDCPGTVLDMIQHYFLTYKATPNHLIKG SPAKIEIVGIYGKKEAQKVIQLAHEDYLSYIGDTAEVN

CPn_0919 1049375 1050430 ldh-Leucine Dehydrogenase FMKYSLNFKEIKIDDYERVIEVTCSKVRLHAIIAIHQTAVGPALGGVRASLYSSFEDACT DALRLARGMTYKAIISNTGTGGGKSVIILPQDAPSLTEDMLRAFGQAVNALEGTYICAED LGVSINDISIVAEETPYVCGIADVSGDPSIYTAHGGFLCIKETAKYLWGSSSLRGKKIAI OG IGSVORRLIQSLFFEGAELYVADVLERAVQDAARLYGATIVPTEEIHALECDIFSPCA RGNVIRKDNLADLNCKAIVGVANNQLEDSSAGMMLHERGILYGPDYLVNAGGLLNVAAAI EGRVYAPKEVLLKVEELPIVLSKLYNQSKTTGKDLVALSDSFVEDKLLAYTS

CPn_0920 1051423 1050431

cysG-Sulfite Synthesis/biphosphate phosphatase
ILEENSMHSELPNYQNIVESVVTEITTQLLNYRSEHRLVPFWEKSDGSFITAADYGSQYY
LKQQLAKAFPNIPFIGEETLYPDQDNEIPEILKFTRLLTSSVSRDDLISTLVPPPSPTS
LFWLVDPIDGTAGFIRHAFAVAISLIYEYRPILSVMACPAYNQTFKLYSAAKGHGLSIV
HSQNLDRRFVYADRKQTKOFCEASLAALNQOHHATRKLSLGLPNTPSPRRVESQYKYALV
AEGAVDFIRYPFIDSPARAWDHVPGAFLVEEAGGRVTDALGAFLEYRKESLVLNNHAVI LASGDQETHETTLAALQNQLNVVPTDKLIAL

CPn_0921 1051526 1052293
snGlycerol-3-P Acyltransferase
GELMLIKLWRATYEGMYTFLVGALLKLRYRMQVEGWDTLNINPKQGCLFLANHVAEVDPI
LLEYLFWSRFHVRPMAVEYLFHSRVVQWFLNSVRSIPIPOLVPCKESKRSLERMNVCYEE
ASRALNRGESLLLYPSGRLSRTGKEEIVNQYSAYVLLHRVMECNVVLVRVSGLWGSAFSR
YKQNSTPKLGPAFKEAFRALLRRGIFFMPKRFVKITLCQVDHLFLKQFPTKQDLNTFLAS WFNOGDDNLPIEVPYA

1052266 1053927 CPn_0922

CPn_0922 1053266 1053927

aas-Acylglycerophosphoethanolamine Acyltransferase

GFAHRSSLRITRKLRMHDORNRGHNNHNLRLRPGSTLLEAFLILGSEHEEGIACFDEHL

GSLSYRELRNAIIAVAIKVSKFSEDRVCVMMPASIGAFIAYFGILLAGKTPVMMNMSGGL

RELRACTKTVEVRRVLTSQOFIKHLTEVQGFVEYPFDLMYMEDVRKRLSWMEKCRIGLYS

KCSVPWLLRIFGVSGVESDDTAVILFTSGTEKLPKAVPLTHKNLMENQEACLKFFDPNTQ

DVMLAFLPPFHAYGFNSCGLFPLLMGVHVVFASNPLNPKKLVEFIDDKKVTFFGSTPVFF

DYILKTAKKQNSCLESLRLVVIGGDALKDTLYEETKKLQPQIALYQGYGATECSPVISIT

TKESPRKSECVGMPIEGMDVLIISKETHIPVSSGEQGLIVVRGNSVFSGYLGNHEHGSFV

SLGGDQWYLTGDLGHIGPSGDLFLEGRLSRFVKIGGEMVSLEALESILHEHFTENQNEDA

GSLVVCGTPGDKVRLCLFTTLATTIHEVNDILKSAETSSIVKISYVHQVESIPILGIGKP

DYVSLNALAVSLFG

CPn_0923 1053966 1055093
b1oF_1-Oxononanoate Synthase_1
vckesflttsDvidfvtNdflgfarsPtiycevskrfQihcQQfphekLGirgsrLmvGpssvtddleskiasyhdapnafivnsGymanlGlchhvSrstdvLLwdeevhmsvvhslsa isQQhhtfhhnnlehLeslLQCYRISSKGRIFIFVSSVYSFRGTLAPLEQIIALSKYHA HLIVDEAHAMGIFGDDGKGLCHALGYENFYAVLVTYGKALGTMGASLLTSSEVKYDLMQN SPPLRYSTSLSPHTLISIGTAYDFLASBGEIARKQVFKLKEHFHECFDSHAPGCVQPIFL PHTCLEEA I SVLETTG I HVGVVAFAKHPFLRVNLHAYNTVDEVNLLAQVMKPYLEKSSHR VH I NHEFHLWRELCCH

CPN_0924 1057301 1055028

PTIA-PTIMOSOMA1 PROCEIN N'
KRFTAKTKSMGYIESSTFRLYAEVIVGSNINKVLDYGVPENLEHITKGTAVTISLRGGKK
VGVIYQIKTTTOCKKILPIIGLSDSEIVLPOOLLDLEMISOYYFAPLGKTLKLFLPAIS
SNYLOPKOHYRVVLKQSKAKTKEILAKLEVLHPSQGAVLKILLQHASPPGLSSLMETAKV
SOSPIHSLEKLGILDIVDAAQLELQEDLLTFFPPAFKDLHPEQOSAIDKIFSSLKTSQFH
THLLFGITGSGKTEIYLRATSEALKQGKSTILLVPEIALTVQTVSLFKARPGKDVGVLHH
KLSDSDKSRTMRQASEGSLRILIGPRSALFCPMKNLGLIIVDEEHDPAYKQTESPCYHA
PDVAVMPGKLAHATVVLGSATPCLESYTNALSGKYVLSRLSSRAAAAHPAKISLINMNLE
PEKSHKHLIFGOPVLKRIAERLEVGEQVLIFFPNRGYHTHVSCTVCKHHLKCPHCOMVLT
FHKYAIF/LLCHLCNSJPKDLPQCCPKCLGTMTLQYRGSGTEKIEKILQOIFFQIRTIRID
DDTTKFYGSHETLLRCPATNIKADVLICTOMIAKGMNPSAVTLAVILKDDSGLYIPDFRAS
SOVPOLITOVAVIBRSRIHLNIELLIGGELDHPTHSAMPDDYSAFYOGETGRELCEYP
PFIRLIPGIFMCKCPKQTWEEAHPVNTLKEQLESTNPLMF/TPCGHFKIKDTFRYQFLI
PSAYVIPVNKKLHHALMLAKLSPRVKFMIDVDFMTTFF

CPN_0925 1057715 1057226 CT779 bypotherical protein PHWLFMENSONFHDTLCOLLD# COFEL CPPLAGLENVTLPHTA LGACYGSTPEKAVEVPN AERQHTTEPPFTNLQEKTKI GUMKCVPHHPDLGONATLÆKYPALFICGILFARKTEGST FVYERIMEEVLEFNRIAKTURGGLEFTFLITLHAKTNIFYNNENFFLALAPLAVTRYKTP TTDYHQSLTQNGCIFLPLYSSLEYFKDSQLKRNLWAILNPLPFAYTPYSS

1058009 1058557 CPn 0926

CPT_UP26 1094009 1094557
Thioredoxin Disulfide Isomerase
CHHTQQTTYDTRFFKDSDMKFWLQGCAFVCCLLLTLPCCAARRRASGENLQQTRPIAAANL
QWESYAEALEHSKQDHKPICLFFTGSDWCMWCIKMQDQILQSSEFKHFAGVHLHMVEVDF
PQKNHQPEEQRQKNQELKAQYKVTGFPELVFIDAEGKQLARMGFEPGGGAAYVSKVKSAL

ODM_0007 43 1059818 1058670

THERE TO REPORT OF THE CONTROL OF THE CONTROL INTERPREDICT OF THE FIRST INTERPREDICT OF THE F SAADRVVIQYDDLVVDSLAIKIPHALPHRWILYSQONSGLMENLFDRGDSSLHQLAKATG SNLLVFNYYPGINSSKGEAKRENLVKSYQACVRYLRDETTGFKANQIIAFGYSLGTSVQAA ALDREVTDGSDGTSWITVKDRGPRSLADVANQICKPIASAIIKLVGWNIDSVKPSERLRC PEIFIYNSNHDQELISDGLFERENCVATPFLELPEVKTSGTKIPIPERDLLHLNPLSPNV VDRLAAVISNYLDSENRKSQOPD

1061035 1059884

*CHLPS 43 kDa protein homolog_3
RRKDFAFTLLNLSNRSDILSGIFSNPHPVSYFSSTHAKQLSDFSKKHPILTKIVTIIVKI FKLLIGLIIPPLGIYMLCQLVCSLALFPRSSMLYSVLKTCFKKYRLGQEIQDYYVKNLDP FFKDPAVSESKRITIQQDHLTIDTLAIHFSTARPKRWLLISLGSGDFLEBMIGLKDSLFL SWKELAKLLGANILIYNYPGVKSSTGKLNLENLATAHNLCAKYLQDKIQGPGANEIITY YSLGGVVQSAALQKNPFTNSETSWVAVKDRAPHSLPAAANSFFGPIGKLIAVLARWKMDA EKNSRELPCPEILVYSADRFRPSEVGDDTALLPEFTLAHAIKRTPFARSKKFIGEVNLLH SSPLKHPTIQKLAEAILESLSRKN

CPn_0929 1062301 1061186

*CHLPS 43 kDa protein homolog_4

EKFWAPIHGSNAFVEDILHSHPSPQATTYFSTRAQKLHEFKDRHFVLTRIASVIIKIFKV
LIGIIILPLGIYWLCOTLCTNSILPSWLLKIFKKQPNTKTLKTNYLHALQDYSSKNRVA

SMRRVPILQDNVLIDTLEICLSQAFTNRWMLISLGSDCSLEËIACKEIFDSWQRFAKLIG
ANILVYNYFGVMSSTGSSSLKDLASAHNICTRYLKDKEQGFGAKEITYGYSLGGLIQAE
ALROQKIVANDDTTWIAVKDRCPLFISFEGFHSCRRIGKLVARLFGWGTKAVERSQDLPC
LEIFLYPTDSLRRSTVRQNKLLAPELTLAHAIKNSPYVQNKEFIEVRLSSDIDPIDSKTR VALATPILKKLS

CPn_0930 1062851 1063330
NO robust homolog present in Genebank/EMBL as of 11/7/98
NKMSELAPCSTGLOMYPHTOVHHALDTRRVILTIAACLSLIAGIVLVGLGAAAILPSLFG
VIGEMILIFSSIALTYLYKKTREVDQIALEPLPEMISKDQSIIDFVKTRDYASLEKKAT
FAYTHTYYDGSMVFYREIPRFMLGSYLALRKDMDRQALF

CPG_0931 1064078 1065718

lys5-Lysyl trna Synthetase
IDERVLGWKSDIYTNILEERNTARAEYLDHEDFLYRSHKLQELSELGVVLYPYEFPGVFS
CEDIKKTFASQELGNSEAAMSRSTPRVRFAGRLVLFRAMGKNAFGQILDHNQTIQVMFNR
CEPISYHGLSEDAEITPIKFIEKKLDLGDILGILGYLFFTHSGELTVLVETVTLLCKSLLS
LPBKHAGLSDKEVRYRKRWLDLISSREVSDTFVKRSYIIKLIRNYMDAHGFLEVETPILQ LPBOTAGLSDASVKYKKUDLISSKEVSDIFVKKSTIKLIKNYMDAHGFLEVETPILQ
NIÝČČÁSEAKPPTTTMEALHSEMFLRISLEIALKKILVGADRIYELGKVFRNEGIDRTHN
PEFFMIERVANYMDYKEVMVFVENLVEHLVRAVNHDNTSLVYSYWKHGPQEVDFKAPWIR
MTMKÉSIATYAGIDVDVHSDOKLKEILKKKTTFPETAFATASKGMLIAALFDELVSDMLI
APBHITDHPVETTPLCKTLRSGDTAFVERFESSCLGKELCNAYSELNDFIRORELLEQOH
TKKELLPDSECHPIDEEFLEALCQGMPPAGGFGIGVDRLVMILTNAASIRDVLYFPVMRR FDAEKTN

CPREB932 1067160 1065721
CYSS-CYSteinyl tRNA Synthetase
VKSDFVMAFSHIEGLYFYNTASQKKELFFPNHTPVRLYTCGPTVYDYAHIGNFRTYVFED
UKSDFVMAFSHIEGLYFYNTASQKKELFFPNHTPVRLYTCGPTVYDYAHIGNFRTYVFED
LKRTLVFFGYSVTHVMNITDVEDKTIAGASKKNIPLQEYTQPYTEAFFEDLDTLNIARA
DFYPHATHYIPOMIQATKLLEQGIAYIGQDASVYFSLNRFPNYGKLSHLDLSSLRCCSR
ISADEYDKENPSDFVLWKAYNPERDGVIYWESPFGKGPFGWHLECSIMAMELLGDSLDIH
AGGVDNIFPHHENEIAQSEALSGKPFARYWLHSEHLLIDGKKMSKSLCMFLTLRDLLHQE
FTGQEVRYMLLOSHYRTQLNFTEEALLACRHALRRLKDFVSRLEGVDLPGESFLPBRTLDS
SSQPFEAFSRALANDLNVSTGFASLFDFVHEINTLIDQGHFSKADSLYILDTLKKVDTVL
GVLETTTSVCIPETVMQLVAEREEARKTKNWAMADTLRDEILAAGFLVEDSKSGPKVKPL

1067532 1068578

CPn_0933 1067532 1068578
predicted disulfide bond isomerase
PVILLQNIKRCSLKQLKVLATLLLSLSLPTLEAAENRDSDSIVWHLDYQEALQKSKEAEL
PLLVIFSGSDWNGPCMKIRKEVLESPEFIKRVQGKFVCVEVEYLKHRPQVENIRQQNLAL
KSKFKINELPCNILLSHEEREIYRIGSFGNETGSNLGDSLCHIVESDSLLRRAFPMMTSL
SLSELQRYYRLAEELSHKEFLKHALELGVRSDDYFFLSEKFRLLVEVGKMDSEECQRIKK
RLLNKDPKNEKQTHFTVALIEFQELAKRSRAGVRQDASQVIAPLESYISOFGQQDKDNLW
RVEMMIAQFYLDSDQWHHALQHAEVAFEAAPNEVRSHISRSLEYIRHQS

CPn_0934 1068948 1068526
rnpA-Ribonuclease P Protein Component
YFVHPLTLPKQSRVLKRKQFLYITRSGFCCRGSQATF//VPSRHPGTCRMGITVSKKFGK
AHERNSFKRVVREVFRHVRHQLPNCQIVVFPKGHKQPPVFSKLLQDFINGIPEGLHRLGK

TKATTGGECTPKSEKCVTAPR

CPn_0935 1069100 1068957 r134-L34 Ribosomal Protein EDTVKRTYQPSKRKRRNSVGFRTRMATRNGRKLLNPFPRHGRHSLVDL

CFn_0936 1069330 1069470 r136-L36 Ribosomal Protein

YLMKVGSSVKADPSKGDKLVRRKGRLYVINKKDPNPYGRQAGPARKK

1069487 1069798

TOTA-STA REBOSOMAL Profesi VKRMAKKOOVAKRAKRREDVEANFKKRODERKIVKOLOVOEEEKENARISLNKMKRDTSP TREAINROLUTORPROYERKPALORIGEROMALMOETPOVIKA W

CPT_0008 1070355 L060840 CT788 hypothetical profeth -{Louder (49) profete periplasmic] QFINICXTHIARMLES ILLEYVILGXCDAY LADKKERIT/COMPANATEGETOLVVLLLL ESHRBALIEKPONIPPONDIEDDILKSCA.KORDETE PSZODLOGIY UTTEKNEYLNKDRENV GPTSPEELVVLLKSKTYPTETWWKKGMKDWORVKDYPSLOOALKEASK

CPn_0939 1070h29 1071175
CT790 hypothetical protein
HinnwrikLSLTLIISTVL/FFSEEIELIGGGKMEKQNLKLDVKEIEFPETVFSRDIETR
VIOVIILHCLAKINGVSLLGGNLIDALFGRDIERMKGIYVEQDSKNHLVKVRVEVNVDYG VSIPEKTEEIQGCIVSEISEYTGLHVAAVHVIIKGLTQPKDRIDEEIEEEVSVQDLPSPE DFLLENSEG

CPT_0940 1073039 1071204

"JET EX HIMELIA WILLIAM INTERVILLA MELLA MEKADOR ERIPFLIMKKTASIETIVVSNETEALLLENNLIKQHHEKYNVLLKDDKTFFCLAISLSHSW REVEARTKAITSSQOLIFOPYVSAEACHTLLEVISQWFPLRTCSDREFALRKRPCILY DMKRCLAPCVGYCTPEEYQGTLDKAILFLKGKIEEVVKDLEKVIQKASDNLEFEQAANYY RTLSLIKQAMAKQQVEKFHFQNIDALGLYRHKQRTILTLLTURSGKLLGARHFSFFENAQ EDQDLLSSFILQYYVSQPY FREILTPPLEFE FTLSYVLNAESPPRISKEBPKTGYGKELLD LAYRNAKAYAATTLPSSTLPYQDFQNILRMSQYPYRIECYDNAHMQGAHATGVYIVFENN GFDPRQYRTFSIDSEKTQNDLALLEEVLLRRFHSLTTALPDMIVVDGCKTHYNKTKKIIQ FTLNILTGUVVTIAKEKSNHSRGLKMEKIFCETFPEGFSLPFTSNLLQFFQILRDEAHRFA ISKHRKRGKALFEQEKIPGIGEVKRKRLLQKFKSWKQVMLSSQEELEAIPGLTKKDIAV LLAROKDFNKSD

CPn_0941 1075504 1073018 mutS-DNA Mismatch Repair

muts-DNA Mismatch Repair
wits-DNA Mismatch Repair
wits-PNA Mismatch Repair
wits-PNA Mismatch Repair
wits-PNA Mismatch Repair
wits-PNA Mismatch Repair
wits-PNA Mismatch Repair
wits-PNA Mismatch Repair
Lipkerpnyiotyvorligkgfkonalogegepakekeskkiopmardiorfotyprotilist
Lipkerpnyioningissergefacldistgsffieecentkelvdeicrlapsevlscnk
fynketaivmoloohlkultistyadwafehkerasoklithhovasldofglkglvpaina
agglisyiodklilptkhiaipotrgkooklidtasovniellaplndpogknsllrim
DHTSTPMGGRLLROILISFTYNPKEILVRODAVEFFLROVTLERNIKTYLCOVRDIERLM
TKVTTGLAGPROIGTLRDSFSAGAQIYEOLASATLPEFFIDKCSLDTKLASLIALLSKSL
MSDLPLRVSDGNIFVDEFHNDLKRLRHNQEHSQEWIWEYQERIRKETGIKKLKIFAQAL
GYYIEVSSEFAPOLPKDFIRROSSRLHAERFTTIELQOFODDMSNISEKLOTLETOFFKDL
CSHILOLRTEILALSOSLADLDYIISLADLAHAQGSYCPHVDMSDTLCIYRCCHPVAKTL
VDTGKFIPNDTEMRGSOTRMILLTGPNMAGKSTYIRQIALLVIMAQMSYIPAKSAHIGV
IDKIFTRIGAGDNLSKGMSTFMVEMAETANILHNATDRSLVILDEVGRGTSTYDGLAIAQ
AVVEYLLFTDKKKAKTLFATHYKELTTLEDHCPHVENFHAGVKDKAGQPVFLYEILKGHS
QKSFGIHVARLAGFPLCVVSRAQQILRQLEGPESITRPAQDKMQQLTLF

CPn_0942 1079 dnaG/priM-DNA Primase 1075955 1077754

dnaG/prim-DNA Primase
NCSITKLRTAMYTEESLDNLRHSIDIVDVLSEHIHLKRSGATYKACCPFHTEKTPSFIVN
PAGAHYHCFGCGAHGDAIGFLMOHLGYSFTEAILVLSKKFQVDLVLQPKDSGYTPPQGLK
EELRHINSEAETFFRYCLYHLPEARHALQYLYHRGFSPDTIDRFHLGYGPEQSLFLQAME
ERKISQEQLHTAGFFGNKWFLFARRIIFPVHDALGHTIGFSARKFLENSQGGYYVTPET
PIFKKSRILIFGLNFSRRIAKEKKVILVEGQADCLQMIDSGFNCTVAAQGTAFTEEHVKE
LSKLGVLKVFLLFDSDEAGNKAALRVGDLCQTAQMSVFVCKLPQGHDPDSFLMQRGSSGL
IALLEQSQDYLTFLISEKMSSYPKFGPREKALLVEEAIRQIKHMGSPILVYEHLKQLASL MMVPEDMYLSLANPQVTAEPQNIPIKQKVPKIHPHIVMETDILRCMLFCGSNTKILLTTAQ FYFVPEDFKHPECRKLFAFMISYYEKYRKNVPFDEACQVLSDSQILQLLTKRRLNTEALD TIFVQSLQKMADRRWREQCKPLSLNQNIQDKKLEILEDYVQLRKDRTIITLLDPESELIP

CPn_0943 1077972 1078238 CT794.1 hypothetical protein FFMKSFKFLIPFLSVILCCONLLSSPRSRAISVTESIGMSAVKTLVLSEKAHEFLEGIGY GVGASSILRDWQTQQWLEIESLLAQNEVM

CPn_0944 1078503 1078997
No robust homolog present in Genebank/EMBL as of 11/7/98
IKIMMHRYFIPLLALLIFSPSLVRAELQPSENRKGGWPTQLSCAEGSQLFCKFEAAYNNA
IEEGKPGILVFFSERPTPEFADLTNGSFSLSTPIAKGFNVVVLCFGLISPLDFFHKMDPV ILYMGSFLEMFPEVEAVSGPRLCYILIDEQGGAQCQAVLPLETKN

CPn_0945 1079001 1079660 CT795 hypothetical protein SIFKNKILPSYFGHNFDQLRRHYMRIALSLLSLLMIFPIFGEESRPGSEDGNSNTQEIVG SQDTQVCLYHSYEQGLQASRIEGRPLVIVVLCNSGDDQQACTIGLSETCEEVLSVLSGSI FSELANFVVLVPSGVNPLIYPPIEDPILAEIVKFKELFKDESFPTGLSIIVVGVTPEGPG DIIEVSPVSLTVEEEETLPSEQTTEVESTSELQSEDPAIA

CPn_0946 1082816 1079745
glyQ-Glycyl tRNA Synthetase
GECQKKKCYTLESFVSEHPLTLQSMIATILRFWSEQGCVIHQGYDLEVGAGTFNPATFLR
ALGPEPYKAAYVEPSRRPQDGRYGVHPNRLQNYHQLQVILKPVPENFLSLYTESLRAIGL
DLRDHDIRFIHDDWENPTIGAWGLGWEVMLNGMEITQLITYFQAIGSKEDLDTISGEITYGI
ERIAMYLQKKISIYDVLWNDTLTYGQITQASEKAWSEYNFDYANTEMWFKHFEDFAEEAL
RTLKNGLSVPANDFVIKASHAFNILDARGTISVTERTRY 1ARIRQLTRLVADSYVEWRAS
LNYPLLSLSSTSEPKETSESVY WHISSTEDLLEIGSEELPATFVPIGIQLESLARQVL
TDHNIVYEGLEVLGSPRRLALLVXNVAPEVVQKAFEKKGPMLTSLFSPDGDVSPQQQFF
ASQGVDISHYQDLSRHASLAIRTVNGSEYLFLLHPEIRLRTADILMQELPLLIQRMKFPK
MWVMDNSGVEYARPIRWLVALYGEHILPITLGTIIASRNSFGHRQLDPRKISISSPQDVV
ETLRQACVVVSQKERRMIIEGGLRAHSSDTISAIPLPRLIEEATFLSEHPFVSCQFSEQ
FCALPFKELLIAEMVNHQKYFFTHETSSGAISNFFIVVCDNSPNDTIIEDNEKALTPRLTD
GEFLFKQDLQTBLTTFIEKLKSVTYFEALGSLYDKVERLKAHQRVFSTFSSLAASEDLDI
AIQYCKADLVSAVVNEFPELQGIMGEYYLKHANLPTASAVAVGEHLRHITMOGKLSTIGT
LLSLLDRLDNLLAGFILGLRFTSSHDPYALRRQSLEVILVJASRLPIDLASLLDRLADH
FPSTIEEKVWDKSKTIHEILEFIWGRLKTFMGSLEFRKDEIAAVLIDSATKNPIEILDTA
EALQLLKEEHTEKLAVITTHNRLKKLLSSLKLSMTSSPIEVLGDRESNFROVLDAFFGF
PKETCAHAFLEYFLGADLSNDIQDFLNTVHIANDDGAIRNLRISLLLTAMDKFSLCHWE
UVAV

1083433 1084059

CPE_0947 1083433 1084053 pdps Glycerol (P Phosphatydylt Landlerase GSRVGLPNYITPURLETFIEM LLYLEGEWEY, PEPVVLEGVILLALLAIGEUFDAIDUVVA PKESO/TDUJKLDOMADSIYRIGIYUEPTOPEVNLEGLD/FIELARDUVLYTLRTVCAF PGRVYAARALKKLKAITLQSVSETILLLUMIEHGLGILLUQNGLETFAUVTVGTTAVYGTAS GIEYFVMMKNILGQRAKTKESEPNHEGPD

1085483 1084047

BOLAXEEPOKKOOREN GÄNEREL. GEGWELAONANELLEAKOREN DAANGESKREVKONDAEAPELEKERTOOAFGE GEGAAN.

YLQEADPADIVHILHDWHYGLLAGLLKNPLNPVHSKIVFTIHNFGYRGYCSTQLLAASQID
DFHLSHYQLPRDPQTSVLMKGALYCSDYITTVSLTYVQEIINDYSDYELHDAILARNSVF
SGIINXIDEDVWNPKTDPALAVQYDASLLSEPOVLTTKKEENRAVLYEKLGISSDYFPLI
CVISRIVEEKGPEFMKEIICHMEHSYAPILIOTSQNEVLLNEFRNLQCLASSPNIRLI
LDFNDPLARLTYAAADMICIPSHREACGLTQLIAMRYGTVPLVRKTGGLADTVIPGVNGF
TFFDTNNFNEFRAMLSNAVTTYRQEPDVWLNLIESGMLRASGLDAMAKHYVNLYQSLLS

1085887 1086483 CPB_0949

PORTON TO THE PROPERTY OF THE PROPERTY OF THE PROPERTY OF THE PROPERTY OF THE PROPERTY OF THE PROPERTY OF THE PROPERTY OF THE PROPERTY OF THE PROPERTY OF THE PROPERTY OF THE PROPERTY OF THE PROPERTY OF THE PROPERTY OF T AGIETITPLKEVAITVSRR

CPn_0950 1086470 1087027
pth-Peptidyl tRNA Hydrolase
PSLEDNMAKLIVAIGNERHGYANTHHNAGFLLADRLVEELQGPPFKPLSKCHALMTLVES
SSGPLVFIKPTTFVNLSGKAVVLAKKYFNVALSHILVLADDVNRSFGKLRLCFNGGSGGH
NGLKSITASLGSNEYWQLRFGVGRPLEEGVELSNFVLGKFSEEENLQLGSIFVEASTLFT EWCSKF

CPn_0951 1087113 1087457 rs6-56 Ribosomal Protein FEYLMGKKENQLYEGAYYFSVTLSEEARRKALDKVISGITNYGGEIHKIHDOGRKKLAYTI RGAREGYYYFIYFSVSPGAITELWKEYHLNEDLLRFMTLRADSVKEVLEFASLPE

CPn_0952 1087469 1087723
rs18-S18 Ribosomal Protein
GEMMNKPVHNNEHRRKRFNKKCPFVSAGWKTIDYKDVETLKKFITERGKVLPRRITGVSS RFOGVLSOAIKRARHLGLLPFVGED

1087727 1088248

rl9-L9 Ribosomal Protein
FKGRRMKQQLLLLEDVDGLGRSGDLITARPGYVRNYLIPKKKAVIAGAGTLRLQAKLKEQ RLIQAAADKADSERIAQALKDIVLEFQVRVDPDNNMYGSVTIADIIAEAAKKNIFLVRKN FPHAHYAIKNLGKKNIPLKLKEEVTATLLVEVTSDNEYVTVLAQGKQTEENQEG

1088259 1088708

ychB-Predicted Kinase
GRKVCYKDIMQYPSPAKINLFLKIWGKRFDNFHELTTLYQAIDFGDTLSLKNSMKDSLSS
NYMELLSPSNLIWKSLEIFRRETQIHQPVSWHLNKSIPLQSGLGGGSSNAATALYALNEH
F@HIPITTLQWAREIGSDVPFFFLQEQH

C₽_0955

CPM_0955 1088612 1089175
(Efame-shift with 0954)
RAFFNPYSYNNIATLGSRINKRCSFFFSSGTALGKGRGEHLFSIKKLNHKHKYVLYLDHQ
GEPEKAVQSLLPQDYSTGNHNACFYGENDLEKSVFRIRTDLKNKKHMLERMMSPFESHV
LMSGSGATLFVCYLEELEQDSKVSSQIHSLIKQTQGIPVSRLYREPHWYSLKQSTYKNSP LECEOPOI

CHRO0956 1089545 1090909
CT805 hypothetical protein
LVMFSMILPPYSYSLKIGAAVLFFCSILHTFLTPWLYTLCQSYEHKKLVFPECWKRYARL
SEJERILSRVEIVFFLMAVPLFFWFLYTEGYRISMAVFNSRNYGFAVFIMVILILLESRP
IVVFFAELVLSSIAKLGKTSPKSWWWTLMIAPPLLSCLLKETGAMIIGATLLMRHFYVFSP
SRRFAYATMGLLFSNISIGGLTSYVSSRALFLIFPALKWEHSFFLSHFAWKAIVALLIST
TIYYFIFRKEFKKFPDIPSDKDPSVEKVPWWIICVNIIFVGSIILSRSTPLFMGALLLFY LCFEKFTIFYODPINLSKVCYVGLFYAGLVVFGDLQEWWVLNLMQGLSDFGYMTVSYTLS IFLDNALVNYLVHNLSVATDCYHYLVVAGCMAAGGLTLVSNIPNIVGYLILRSAFPSSTI HMCWLFLGALGPSIISLGVFWLLKNVPEFLYCFFR

CFT=0957 1093812 1090963
ide/ptr-Insulnase family/Protease III
KIYTRNCKMFWKLLCPILICTSLSITSCEQQFKVVPNQCPLQVSTPAAADQKIEKIICSN
GLPŁLIISDPNDTSGAALLVKTGNNADPEEYPGMAHFTEHCVFLGNEKYPEVSGFFGFL
SENWSVHNAFTYPNKTVFVFSVEHSAFSDALDQFVHLFINPKFRQEDLDREKYAVHQEFA
AHRLSDGRRVHRIQQLVAPQGHPCARFGCGNASTLTPVTTEKMAEWFKLHYSPENMCAIA
YTŠÆLSKAKKQFSKIFSQIPRSKNYERQEPFLPSGTTSSLKNIYINQAIQPTSNLEIYW
HIYESSHPIPLGCYKALAEVLRNESKNSLVSLLKNEQLITDLDVEFFRSSLNTGEFYISY
ELTEKGDKHYSQVIDSTFQYLRYIQEHGIPNYTLEEISTINALNYCYSSKSPLFDLLCKQ
IVSLGNEDLSTYPYHSLVYPKYSSEDESALLNLVSDPEQARFVLSSKNSEHWEEPGAPAIS IVSIGNEDLSTYPYHSLVYPKYSSEDESALLNLVSDPEQARFVLSKNSEHWEEATQLHD
PIFDMTYYVKALDGVODYGKVOSLKPIALDFKPNLFIPKEVTLDGVHLLKKQEFPFAPALS
YODDKLTLYHCEDHYYTAPKLSSQIRIRSPQISRSSPOFLVATELYCLAVNDQLLREYYP
ATQACLSFTSALGGDGIDLRVSGYTTTTVPALLNSILTSLPNLEISYEFFLVYKKQLLELY
QGALLNCPVRSGLDELASQVMKETYSNTTKLSALEKLSFSEFQAFASNLFNSVHLEVMVL
GNLSEQOKKDYLEMLQVFTASRSSHATKPFYYELQSGEISEIHHDYPLTANGMLLLLQDK
SSPSIQGKVCAEMLFEWLHHITFEELRTQQLGYMVGARYREFASRPFGFLYIRSDAYSP
EELLAKTSLFLNKVSASPEKFGISQEKFANIRKAYINKILEPEHSLDMMNSALFSLAFER
PFVEFSTPDLKIAIAETLTYEEFLKYCQCFLSNELGTQTSVYIRGTQKTS

1094803 1093793

plsB-Glycerol-3-P Acyltransferase IYRAIYMQFSRYLRYAFDNQYLPEPLYQKFSVFHQNYIDAATKKAAADQAEVLCLQWVKV II EDLKNPF IFPPYHKKIRAPIDLFRLSIDFFSLVI DDKNSRILLNIHRLKE IESYLARD NVVLLANHOTECDPQLMYYALGKTHPELMENMI FVAGDRVTSDPLARPFSMGCDLLCIYS KRHIATPPELREEKLLHNOKSMQILKTLLNEGGKFIYVAPACGRDRKNAEGRLYPSEFSP ESIEVFRLLAKASNQTTHFYPFALKTYDILPPPPKIENAIGEQRAIFFAPVFFNFGAELF FDALCCKEELIHCDKHAQRTLRAEKVFSIVKNLYEEL

1096376 1094799

CHE-0757

GATE-AXIAL FILAMENT PROTEIN
ACGYGICTRKVMENEILLNIESKEIRYAHLKNGOLFDLTIERKKVPOLKGNIYRGRVTNI
LRNIGGAFINIDERENGFIHEGDILENGKFEOMFOMDVDALPEBASEAPLLGSEBAPIE
EFILKELDFVLVGVVVKEFIGGRGARTENICI FORGYLVLLPNSPHRGVGRKIEDFIMREQL
KQLIKGFEMFOFMKLIGRTAGTTAGTEALINEAHDLLLTWKTILEKFYSTEOPCLLYSET AGEILANDAMENTALI ENDYATYOKEKHMILKYOTRASI KIETYYSTE GWELLTSET DILLKAYITTO I DKNYKREL I EDYATYOKEKHMILKYOTRASI KIETYYSTE GWELLTSET KELDKAYTRKI IMERSCRYLFFOKTEAMHT I DVNSGPSTOLESGVEETLVOINLEAAEEIA RESILMOYOLYTI I ET TENTEN TORKAKTORTUERLKEHMYYDAARCT LLSMSEFGLVEMTROR RESILMOYOLFTLOPYCSGNA I I KITTESVVI E LERDLIKY INHKEHSHLCLVVHPEIASYM KQENDDIRMINLAKOLKAKLQINTSDSVHLMHYQFFSLITYSES I DL

CPn_9960 1996575 10 07102 CT909 hypothetical profein subjustionedkalvidskeeskelvouriklyterikungdterigysispeyirekke eellingfleveosiceridsoowiljusiktoligicepvcnnffshsvclpolorvishde vosgyfdcrplirqelllesdcfeecsoggcperknilkfledrkkhegnspfeyl

1097106 1097297 CPn 0951

rl32-L32 Ribosomal Profess THE STATE OF THE S

CPn_0962 1097301 1098275
plsX-FA/Phospholipid Synthesis Protein
ILSDFMEVQIGIDLMCGDHSPLVVWQVLVDVLKSQSSTIPFAFTLFASEEIRKQIQEEFI
SDLPQEKFPKIISAENFVANEDSPLAAIRKKSSSMALGLDYLQEDKLDAFISTONTGALV
TLARAKIPLFPAVSRPALLVCVPTMRGHAVILDVGANISVKPEEMVGFARMGLAYRQCLG
DSKIPTIGLLNIGSEERKGTEAHRQTFMLRETFGEAFLGNIESGAVFDGAADIVVTDGF
TGNIFLKTAEGVFEFLQRILGDKLEADIQRRLDYTFYPGSVVCGLSKLVIKCHGKACGSS
LFHGILGSINLAQARLCKRILSNLI

CPN_0963

D098374 1103224

DND_21-Putative Outer Membrane Protein

TPLRFKVAMVAKKTVRSYRSSFSHSVIVAILSAGIAFEAHSLHSSELDLGVFNKQFEEHS
AHVEEACTSVLKGSDEVNPSOKESEKVLYTOVPLTVGSSGESLDLADANFLEHFOHLFEE

TTVFGIDOKLWMSDLDTRNFSQPTQEPDTSNAVSEKISSDTKENRKDLETEDPSKKSGLK
EVSSDLPKSPETAVAATSEDLEISENISARDPLQGLAFFYKNTSSQSISEKDSSFQQIIF
SGSGANSGLGFENLKAPKSGAAVYSDRDIVFENLVKGLSFISCESLEDGSAAGVNIVVTH
CGDVTLTDCATGLDLEALRLVKDFSRGGAVFTARNHEVONNLAGGILSVVGNKGAIVVEK
NSAEKSNGGAFACGSFVYSNNENTALWKENQALSGGAISSASDDIQGNCSAIEFSGNQS
LIALGEHIGLTDFVGGALAAQGTLTLRNNAVVQCVKNTSKTHGGAILLAGTVDLNETISE
VAFKONTAALTGGALSANDKVIIANNFGEILFEQNEVRNHGGAIYCGCRSNPKLEQKDSG
ENINIIGNSGAITFLKNKASVLEVMTQAEDVAGGGALMGHNVLLDSNSGNICFIGNIGGS
TFWIGGYVGGGAILSTORVTISNNSGDVVFKROKGCCLAGKYVAPQETAPVESDASSTNK
DEKSLNACSHGDHYPPKTVEEEVPPSLLEEHPVVSSTDIRGGGALAGHIFITDNTGNLR
FSGNLGGGEESSTVGDLAIVGGGALLAESVNITDNGSAVSFSNNTRILGGAGVAAPQGSV FSGNLGGGESSTVGDLAIVGGALLSTNEVNVCSNQNVVFSDNVTSNCCDSGAILAKK VDISANHSVEFVSNGSGKFGGAVCALNESVNITDNGSAVSFSKNRTRLGGAGVAARQGSV TLCGNÇGNIAFKENFVFGSENQRSGGAIIANSSVNIQDNAGDILFVSNSTGSYGGAIFV GSLVASEGSNPRTLTITGNSGDILFAKNSTQTAASLSEKDSFGGGAIYTQNLKIVKNAGN VSFYGNRAPSGAGVQIADGGTVCLEAFGGDILFEGNINFDGSFNAIHLCGNDSKIVELSA VQDKNIIFDOAITYEENTIRGLPDKUVSPLSAPSLIFNSKPQDDSAQHHEGTIRFSRGVS KIPQIAAIQEGTLALSQNAELWLAGLKQETGSSIVLSAGSILRIFDSQVDSSAPLPTENK EETLVSAGVQINMSSPTPNKDKAUDTPVLADIISITVDLSSFVPEQDGTLPLPPEIIIPK CTKLHSNAIDLKIIDPTNVGYEHNALLSSHKDIPLISLKTAEGMTGTFTADASLSNIKID VSLPSITPATYGHTGVWSESKMEDGRLVVGWQPTGYKLNPEKQGALVLNNLWSHYTDLRA LKQEIFAHHTIAQRMELDFSTNVWGSGLGVVEDCONIGEFDGFKHHLTGYALGLDTQLVE DFLIGGSFSFFCKTESGSYXAKNDVKSYMGAYAGILAGPWLIKGAFVYGNINNDLTTD YGTLGISTGSWIGKGFIAGTSIDYRYIVNPRRFISAIVSTVVFVEAEYVRIDLPEISEQ GKEVKTFQKTRFENVAIPFGFALEHAYSRGSRAEVNSVQLAYVFDVYRKGFVSLTILKDA AYSWKSYGVDIPCKAWKARLSNNTEWNSYLSTYLAFNYEWREDLIAYDFNGGIRIIF

CPn_0964 1104812 1103301

No robust homolog present in Genebank/EMBL as of 11/7/98
QSILESIIKYFYLIHNSKMHMSNPISLFSPAELIAKYNLIPKTSPIYPRRTELIILEENA
CQTRLITNVAQVLHPSSLFSMSKKILNPCGCSGGPLCWVILNILAFIITSVLFIILLPVNL
IVAGLRIEMPLEPKKIVEDLSEPTTEETNEVIOPFIFALQALIFEDNKLRSFKIVEGSVG
KAPLPNPFLNRLVAISPQESQEAMRKIPDLCSQLKKVLKSLGVLTPEWKHMLKYFEGLKN
EHDSNPDKKTFPILIKLLIEALTGKSSLPKTPSTKEKMQAALFIASSCKTCKPTWGEVIT ERISHPERTI LITALET EN STORE THE ASSETT ASSET

CPn_0965 1106769 1104925

1pxB-Lipid A Disaccharide Synthase
KGFSFSKVGLANIPSGLVYLLYPLGFLASLFFGSAFSIQWLLSKKRKEVYAPRSFWILSS
IGATLMIVHGTIQSQFPVTVLHVINLIIYLRNLNITSSRPISFRATLVLMALSVVFVTLP
FLYVMBWMASSNIFHLPLFPAQLSWHLIGCLGLAIFSGRFLIGWFYIESNNTKDFPLLF
WKIGLLGGLLALVYFIRIGDPINILCYGCGLFPSIANLRLFYKEQRSTPYLDTHCFLSAG
EASGDILGCKLIQSIKSLYPNIRFWGVGGPAWRQGGLOP ILMMEEFGVGSFABVLGSLFR
LYRNYRKILKTILKHKPATLIFIDFPDFHLLLIKKLRKHGYRGKIIHYVCPSIWAWRPKR
KRILEQHLDMLLLILFFEGLFKNTSLETTVYLGHPLVEEISDYKEQASWKEKFLNSDRPI
VAAFPGSBRGDISRNIRIOWOAFINSSLSOTHOFVVSSSSKYYDEIIEDTLKAEGCOHSO VAAFPGSRRGDISRNLRIQVQAFLNSSLSQTHQFVVSSSSAKYDEIIEDTLKAEGCQHSQ IIPMNFRYELMRSCDCALAKCGTIVLETALNQTPTIVMCRLRPFDTFLAKYIFKILLPAY SLPNIIMNSVIFPEFIGGKKDFHPEEIATALDLLNQHGSKEKQKEDCRKLCKVMTTGQIA SEEFLKRIFDTLPAV

CPn_0966 1108055 1106748

pcnB_2-PolyA Polymerase
LLITIIMVCENNILSGRGLELLKKKSNITLTPTIYSVSNHNIKLKDFSPHALSVIKTLRK
AGYTAYIVGGCTRDLLLNTTPKDFDISTSAKPEEIKAIFKNCILVGKRFRLAHIRFSKQI
IEVSTFRSGSTDEDVLITKDNLWGTPEEDVLRRDFTINGLFYDPEHEEIIDYTGGVNDLR
RRYLRTIGDPFTRFKQDFVRMLRLLKILSRSPFTVETQTOEALIACROELIKSSQARVFE
ELIKMLNSGAARNFFQLLIENHLLEILFPYMDKAFRLNPALEEQTATYLKALDDKILKKE AEYDRHQLMAIFLFPLVNFMVRYKHQKHPYLSLTSVFDYIKNFLEQFFADSFTSCSKKNF ILTALILQMQYRLTPLIFTKKALFFNKKLLHHTRFLEALSLLEIRSIVYPKLDKVYVAWI RHHQTLKCKKDSHSQK

CPn_0967 1108431 1109885 mrsA/pgm-Phosphodiucomutade FTAYKFAFICACRSEKIRRIGIDFRRNMQQSVPYLFGTDGVRGRANUEPMTVETTVLLGK PTAYKFAFTCACGER TRRIGIDFRRIMOQSVPYLEGTIDJVRGRANFERMIVETTVLLGK AVARVLREGRSGKHRVAVGKDFPLSGYMFENALIAGLINDMGLETLIVLGP LEFTFVAFTTR AYRADASIMISSASHNEYNDNGIR TESLEGERK IDJVLEQP LETMYSSADFOPDLEDBHAVGK NKRVIDAMGRYVEFVR NFFFRGPFLKGLK IVLLCAHGACJXKVARSVFEELDARVICYGCE PTG INTNEHGSALFOVIÇKAV IFRIQAHLIGI ALLD IDDDF I IMVDEKGHLVTURMILISICA GDLKKRSALFINRVAVTINTIN GVILK ZIDELBURGVETVYDDHIVLHAMLEHEVTLGXEQ SGHMLFLDYNTVDXILVSTALGACJXLRIMITESESMLCDDITAFTVKGPQYTLINVAVRTK I PLET IPLI ERTURDVQDALAI SCHILLPYGGYPENIC PYMVECHKKHQVDXTAPALADV IDAELG GOSBE TOTRE

404884

CPn_0968 1109889 1111721 glmS Olucosamino frustose & P Aminotransferase

DRMCGTFGYLGNQDGVSTVLEGLAKLEYRGYDSAGLAAVVEQELFTRKTVGRVQELSNLF QEREIFTASVICHTRWATHGVPTEINAHPHVDECPSCAVVNNGIIENFKELRRELTAGGI SFASDTDSEIIVQLFSLYYQESQDLVFSFCQTLAQLRGSVACALIHKDHPHTILCASQES PLIUGLGKESTFIASDSRAFFKYTRHSQALASGEFAIVSQCKEPEVYNLELKKHKDVRQ LTCSEDASDKSGYGYYMLKEIYDQPEVLEGLIQKHMDEEGHILSEFLSDVPIKSFKEITI PRNLAKSVTVE

CPn_0969 1111803 1112999

tyrP_1-Tyrosine Transport_1

VYVMSNKVLGGSLLIAGSAIGAGVLAVPVLTAKGGFFPATFLYIVSWLFSMASGLCLLEV
MTWHKESKNPVNMLSMAESILGHVGKISICLVVLFLFYSLLIAYFCEGGNILCRVFNCQN
LGISWIRHLGPLGFALIMGPIIMAGTKVIDYCNRFFMEGLTVAFGIFCALGFLKIQPSFL
VRSSWLTTINAFPVFFLAFGFQSIIPTLYYYMDKKVGDVKKAILIGTLIPLVLYVLWEVV
VLGAVSLPILSQAKIGGYTAVEALKQAHRSWAFYIAGELFGFFALVSSFVGVALGVMDFL ADGLKWNKKSHPFSIFFLTFIIPLAWAVCYPEIVLTCLKYAGGFGAAVIIGVFPTLIVWKGRYGKQHHREKQLVPGGKFALFLMFLLIVINVVSIYHEL

CPn_0970 1113452 1114648

tyrP_2-Tyrosine Transport_2

YYVMSNKYLGGSLLIAGSAIGAGVLAVPVLTAKGGFFPATFLYIVSWLFSMASGLCLLEV
MTWMKESKNPVNMLSMAESILGHVGKISICLVYLFLFYSLLIAYFCEGGNILCRVFNCON
LGISWIRHLGPLGFAILMGPIIMACTKVIDYCNFFFMFGLTVAFGIFCALGFLKIQPSFL
VRSSWLTTINAFPVFFLAFGFQSIIPTLYYYMDKVGDVKKAILIGTLIPLVLYVLWEVV
VLGAVSLPILSQAKIGGYTAVEALKQAHRSWAFYIAGELFGFFALVSSFVGVALGWMDFL
ADGLKWNKKSHPPSIFFLTFIIPLAWAVCYPEIVLTCLKYAGGFGAAVIIGVFPTLIVWK GRYGKQHHREKQLVPGGKFALFLMFLLIVINVVSIYHEL

1114693 1115415 yccA-Transport Permease

ycca-trainyold felimese EGSMGLYDRDYIQDSRVQGTFASRVYGWMTAGLIVTSCVALGLYFSGLYRSLFSFWWWC FATLGVSFFINSKIQTLSVSAVGGLFLLYSTLEGWFFGTLLPVYAAQYGGGVTWAAFGSA ALVFGLAAVYGAFTKSDLTKISKIMTFALIGLLLVTLVFAVVSMFVSMPLIYLLICYLGL VIFVGLTAADAQAIRRISSTIGDNNTLSYKLSLMFALKMYCNVIMYMYLLQIFSSSGNR

CPQ_0972 1116377 1115430

ftsY=Cell Division Protein Ftsy
RCINNSLIFPSYLVSFLLLQLTLLLAMFKFFRNKLQSLFKKNISLDLIEDAESLFYEADF CTEMPELLCARLRATKKADASTIKDLITYLLRESLEGLPSOASOSSOATRIVSLLLOTING SGKFTTAKLAHYYKERSESVMLVATDTFRAGMDOARLWANELGGFVSGOPGGDAAAI AFDĞTQSALARGYSRVIIDTSGRLHVHGNLMKELSKIVSVCGKALEGAPHEIFMTVDSTL GNNATEQVVVFHDVVPLSGLIFTKVDGSAKGGTLFQIAKRLKIPTKFIGYGESLKDLMEF DLDÜFLNKLFPEVEKI

CPn_0973

CPn 10973 1116346 1117527
"succinyl-CoA Synthetase, Beta"
EGKSKELFMHLHEYQAKDLLASYDVPIPPYWVVSSEEEGELLITKSGLDSAVVKVQVHAG GRÖRRIGGVIVAKSSAGILQAVAKLIGMHTSNOTADGFLPVEKVLISPLVAIQREYYYAV
IMBRHRCPVLMLSKAGGMDIEEVAHSSPEQILTLPLTSYGHIYSVQLRQATKFMEWEGE
WHREGVOLIKKLAKCFYENDVSLLEINPLVLTLEGELLVLDSKITIDDNALYRHPNLEVL
YDPSQENVRDVLAKQIGLSYIALSGNIGCIVNGAGLAMSTLDILKLHGGNAANFLDVGG ASQKQIQEAVSLVLSDESVKVLFINIFGGIMDCSVVASGLVAVMETRDQVVPTVIRLEGT NVELGKEIVQQSGIPCQFVSSMEEGARRAVELSM

"sucD-Succinyl-CoA Synthetase, Alpha"
"sucD-Succinyl-CoA Synthetase, Alpha"
VCRERRYMFHSLSKNTFLITGGITCKAGSFHTEQCLAYGTNFVGGVTPGKGGTLWLDLPV
YDSVLEAKQATGCRATMIFVPPPYAAEAILLEAEEAGIELIVCITEGIPVRDMLEVARVMD
NSTSSELIGPNCPGIIKPGECKIGIMPGYIHLPGNIGVVSRSGSTLTYEAVWQLTQLKIQCS
ICVGLGGDPLNGTSFIDVLQALEEDPYTELILMIGEIGGSAEEEAAAWIQAHCTKPVVAF
IACAMAPKGKRMGHAGAIISGNSGDAKSKIQVLRESGVTVVESPAHIGKTVDAVLRAKEL

1119038 1119637

No robust homolog present in Genebank/EMBL as of 11/7/98
GIEEQVALSIARKILKIILALILFPLVLLAWVIRYQLHANFHCSVVPFROFSVMQAYKCS
EAKIEEMLDLLDLETLEWSSRCLRQOMTFANRLEEELIGGLERVSETEELISLGGKRNLVR
LLLTHFFNPPKKSRVESVGHEVVFPVFDRLKREEEIIGDGPITRSNEELWALLDHGTARG IHKTLWFSIFFKYLTQIELF

CPn_0976 1120079 1121185

No robust homolog present in Genebank/EMBL as of 11/7/98

ILMLDYGCFDFSVPTSPEHRIMAALDRWFFLGGHRAR ILTLEGNHYRAFQENMSISTVEKI

LKLISYLLIPIVLIALLIRGELHSFRKCMWKCDSLSDARVPHDVQPFNDFQLFNNGRELN

IWKNRRYVSGIDVLMVPVDYLRSQFPGFKEIPEAIRCENYVSDQFSEESKTSYLRAMLT

DIVGYILSLDETYWTNVILKIRAMCITFESFFCKEADPNYSPRVTHHYDESWKALARHV

LGECMMVNRLDEALIRTEKPCKEGECITKQFLKDYCKKHLEVMSCPDFIESLVDEKIREF

RCPSILNSAVCDVIDRKCQEHLLKAIINEANRRLFGMKNSSFTMRGNQVLFYTIFSPPKL

PPAASSVF PPAASSVYF

CPn_0977 1121329 1122402
No robust homolog present in Genebank/EMBL as of 11/7/98
LYINGFANTLKSSFLMEVYSFSPSVRTSFOHRVMAALDNWFFLGGRRLKVVSLDSCNSGQ
ACEEVVPISTTERVLKILSYLLIPIVIIALLIPILLIPILLIPITAKVSGKPMLKTLQLGIDIK
SFILEGGHVNTMDSATLFKAIRLEGKRVDVEYHRLHSSDKVVFYIPAQKLPDDLRLTHWL
PEKETRKTEYVRHMLAHVMGYLTSGCKERLQOVVDSRSSTSLAGEKVLGYRFIDHEGSG
GEFGRLLNENITTKGSEDKEVVGSDLFDMAFQCMVPOFISVIQGPTFSEELVHEMSGKLD
LDCIYPEDDEFEGKFLNTLLKAVLHHGFEGISVASMGVIFLICPDSLALQIPFLRNQK

1122654 1123693

CPn_0978 1123634 1123635 No robust homolog present in Genebank/EMBL as of 11/7/98 KYFFMEVYSPHPAVRTSTQHRVMAALDAWFFLGSHRLPVVSLDSTNSGWAYQELVSISTT EKVLKLESVILVPIVIIALLIRGLEHSNFRIDVEKEPWLKIRELGIDIESCKLPSSYVNO VRSFIWFEKDKSKRPRIDVDYHTLHSKDWVFFIVFOKIPKTSPSYWFSQKETRKRDYV VRMIDHVIGYLTHSKGEWLQVISKTSYGSATVLDPERVLQVCTUTDNGELGGEVGRLUMEE SATKSSGOKEVLLSHVSDIICQCWWPKFLEVIQSPAFIEELVEEVSGKLNLDFLCLEKAN "THISGELRNSLERAVVHHCSBSVDIKKVGAGLIIVTEAIGGIPPSRS

1123971 1125443 CPn 0979

htrA-DO Serine Protease GIDMITKQLRSWLAVLVGSGLLALPLSGQAVGKKESRVSELPODVLLKEISGGFSKVATK ATPAVVYLESFEKSQAVTHESPGRRGPYENPFDYFNDEFFNRFGLPSQREKQSKEAVR ATPAVVYLESFEKSQAVTHESPGRRGPYENPFDYFNDEFFNRFGLPSQREKQSKEAVR GTGFLVSPDGYIVTNNHVVEDTGKIHVTLHDQGKYPATVIGLDPKTDLAVIKIKSQNLPY LSFGNSDHLKVGDWAIAIGNPFGLQATVTVGVISAKGRNQLHIADFEDFIQTDAAINPGN GG7ELBILDGGVTGVWTAIVSGSGGYIGIGFAIPSLMANRIIDDLIRDGOVTRGFLGVTL THE ACTION PERSONS AS COUNTY OF NEW YORK OF THE PROPERTY OF TH TKGILIISVEPGSVAASSGIAPGQLILAVNRQKVSSIEDLNRTLKDSNNENILLMVSQGD VIRFIALKPEE

1126988 1125504

*similarity to Saccharomyces serevisiae hypothetical 52.9KD

FVMLNHAKKHAKPYVLIFFSTKDKLSYCDIIFNNCSGKPMNLDSKHFDINSANFLEEFAK FYRELINANGARAPTYLIF STADRISTOLIFF IN SGRPRING BRIT DINSANGLEEF AR
FISFPSISADSDHLODCENCAHELUDHYNK I FÜVELWET PÖHPPI I YASYKSEDPLSFTL
MLYNHYDVOPAQLSDOWKGDPFILREENSONLYARGASDNKOOFFTLKALOHYYESOONF
PLNIIWLIEGEEESGSLALFTWLEKKKEALRADYLLIVDOGFLSEKHPYVSIGARGIVSM
KISLEEGNKDMHSGVLOGIAYNTYRALSEILSSLHHPDNSIAIEGFYDDLALFSDSDRD LPKSDTLRECEENLGFRPQGYEASYSPEESALRPTVEINGISGGYTGPGFKTVIPYRATA YLSCRLVPNQDPDKAAHQVIHHLKQQVPSSLKFSYEILPGGSRGWRSSANLPIVKVLQEI YSDLYNEECLRLVMPATIPIGPLLGEAAQTSPIICGTSYLSDDIHAAEEHFSMDQLKKGF LSICOLLDKLPKIKE

CPn_0981 1127019 1129952

Zinc Metalloprotease (insulinase family)
VTESMKAGDTYRNFIIKSCKDLPEIESKLLEAEHKPTGASIMMIVNNDEENVFNICFRTC
PQTSKVAHVLEHMVLCGSENYPVRDPFFSMTRRSLNTFINAFTGPDFTCYPAASQIPED
FYNLLSVYIDAVFHPLLTKOSFLQDEAWRYEFNSENHLCYTGAVVENMEKGAMMSGEARLSE
ALNAAIFPSVTYGVNSGGEPREIVTLSHEDVRAFHQSQYSINRCLFYFYGNIKPSRHLDF
LEEKLLRQATKLEKQAVSVPLQKEFKEPVRNILTYPVDHQEEDKVLFGISWLTCSILEQQ
ELLALHVLEIILMGTDASPLKSRLIKSGFCKQTEMSIENDIREIFMTLVCKGCSPAGAQK
LEALIFASLEEIIREGISENIVEGAVHQLELSRKEITGYSLPYGLSIFFRSGLIKQHGGS
AEDGLRIHSLFSELRNSLKNSDYLAKLIRKYFLDNPHFARVILLPDTELVAKDNKDEQQL
LLSVSEKLTDENKEKIQQNVRELTESQEQKEDLNGILPNLALDKVPTSGKEFPLIKEGLS
QGEVLHHECFTNDIVFIDVVLDIPPLSGEELPWLRLLVFLMLQLGCGGRSYKEHLEFLLE
HTGGYDVSYDFSPHANKNSFLSPSVSIRGKALSSKSEKLCGIVSDMLTSVDFTDIPRIRE
LLMQHNEALTNSVRNSPMSYAVSMACSGNSITGAMSYLTTGLPYVKKIRELTKNFDQNID
EAWVILQRLYTKCFSGKRQIVISGSAHNYQQLKDNKFYGLLDYLIVIPEPWENPSINLYV EAVVILORLYTKCFSGKRQIVISSSAINYQQLKDNFYGLLDYLIVIPEPWENPSINLYV
TSRGLHIPARAAFNALAFPIGDIAYDHPDAAALTVAAEILDNVVLHTKIREQGGAYGGG
AANLSRGSFYCYSYRDPEIATTYKTFLKGVSELIASONFTKEDIVEGALGVVQGLDMPVAP
GSRASVAFYRLKSGRIPVLRQAFRRSVLEVTKEHICMVMDKYLESTVQETTLISFAGEEM LRNNVLTLDKDFPIVPAI

CPn 0982 1131215 1129962

yigh family KKELASVMNLPVSLACLLLSGCVFFLGVFVSSSLYARKKRAFLEKIQKLEHENQLLQTSL NLSRHQEQLIEDFSNRLALSSHKLIKDMKEEAQNYFGDTSKSFQSILSPIQTTLTTFKQS LETFETKHAEDRGRLKEQISQLLAVEKKLEHETHVLTDILKHPGSRGRWGEIQLERILEL AGMLKYCDYDSQTTSAQGAFRADIIIRLPQDRCLIIDAKAPISDSYFSVEEIDKGDLVDK IKEHIKTLKSKSYWEKFHQSPEYVILFLPGESLFNDAIRLAPELMEIGASSNVILSSPLT LLALLKTIAYMWKQENLQKQIQEVSLLGKELHRRLQVVFTHFQKIGKNLNQTVQSYNDMT SSFQYRVLPTLRKFEGLETSSSHQIEEPTPIESLATSFPHTCDIDTNLAVIESLEKQD

CPn_0983 1132045 1131206
pssA-Glycerol-Serine Phosphatidyltransferase
kNPLCYECKKLWQIDMAGLDLEARGKRRVVTPNATTAFGLCCGLFIIFKSVLRTSSSVEL
FHRLQGLSLLLISAMIADFSDGAIARIMKAESAFGAQFDSLSDAVTFGIAPPLIAIKSLD
GIYVGNFFSSLLLITSIIVSLCGVLRLVRYNLFSQKTVDVSKPYCFIGLPIPAAAASIVS
LALFLASDFFPDLPAQLRVGLLSFALLFIGGLMISPWKFPGVKHFRFNVSSFLLVVTIGL AACLFFSGLVDHFVEVFFLVSWLYTLVGFPIFSIIYRKKS

1132370 1135510

CPn_0984 1132370 1135510

"nrdA-Ribonucleoside Reductase, Large Chain"

(KVMYeVEEKHYTIVKRNGMFVPFNQDRIFQALEAAFRDTRSLETSSPLPKDLEESIAQI
THKVVKEVLAKISEQQVVTVERIODLVESQLYISGLQDVARDYIVYRDQRKAERGNSSSI
IAIIRRDGSAKFNPMKISAALEKAFRATLQINGMTPPATLSEINDLITLRIVEDVLSLHG
EEAINLEEIQDIVEKQLMVAGYYDVAKNYILYREARARARANKDQDGQEEFVPQEETYVV
QKEDCTTYLLRKTDLEKRFSWACKRFPKTTDSQLLADMAFMNLVSGIKEDEVYTTACIMAA
RANIEREPDYAFIAAELLTSSLYEETLCCSSQDPNLSEIHKKHFKEYILNGEEYRLNPQL
KDYDLDALSEVLDLSRDQOFSYMGVQNLYDBYFNLHEGRRLETAQIFWNRVSMGLALNEG
EQKNEWALTFYNLLSTFRYTPATFTLFNSGMRHSQLSSCYLSTVXDDLSHIYKVISDNAL
LSKWAGGIGNDWTDVRATGAVIKGTNGKSQGVIPFIKVANDTAIAVNQGGKRKGAMCVYL
ENWHLDYEDFLELRKNTGDERRRTHDINTASWIPDLFFKFLEKKGMWTLFSPDDVPGGHE
AYGLEFEKLYEEYERKVESGEIRLYKKVEAEVLWRRMLSMLYETGHPWITFKDPSNIRSN
QDHVGVVRCSNLCTEILLNCSESETAVCNLGSINLVEHIPNDKLDEEKLKETISIAIRIL
DNVIDLNFYPTPEAKQANLTHRAVGLOVMGFQDVLYELNISYASQEAVEFSDECSSIIAY ODHYGVNEYDELTELLUSESETAVINUS INDHEHENDKLDEEKIKETISTAIRIL
DNVIDLNEYPTPEAKQANLTHRAVGLGVMGFODVLVELNITGYASGEAVEFSDESSETIAY
YAILASSLLAKERGTYASYSGSKWDRGYLPLDTIELLKETRGEHNVLVDTSSKKDWTPVR
DTIOKYGMRNSGVMAIAPTATISNIIGVTQSIEPMYKHLFVKSNLSGEFTIPNTYLIKKL
KELGLWDABMLDDLKYFDGSLLEIERIPNHLKKLFLTAFEIEPEWIIECTSRRQKWIDMG
VSLNLYLAEPDCKKLSNMYLTAWKKGLYTTYYLRSQAATSYEKSFIDINKRGIQPRWMKN
KSASTSIVVERKTTPVCSMEEGCESCQ

CPn_0985 1135432 1136571
"ntdB-Pidonucleoside Reductade, Small Chain"
ISVHKY7GRKKNNPRLFNGRRERILSITERRGAKMEADILL/SKLKRVEVSKKGLVNCNQV
DVNQLVPIRYKMAWEHYLNGCANNMLPPEVPNAPDIELMYTDELSEDEDERRVILLNLGFFS
"AESELVGNNIVLAIFHHITMPEAROYLLRQAFERAVHTHTPLYICEGLGLDEGEVFNAYN
ERASIPAKDDFOMTLIVDVLDENFSVO7GEGLG/FIKNL/GTYIIMDLIFFYSGFYMILS
FHRONYMMTIGEQYOYILRDTHIHLNTYIDLINGIKEENFGTWHTELGELTVALIEKAVE
LEIEYAPDLEPRGILGHGTMFIDVYHHADRELERIGHEFIYHDLRKAVE EKNETETRVTEYOTAGNIGU

CPN_0986 T1367L2 T137305,
YORK PROTECTED TRINA MOTBY LEGA.
ELLEMKPODILIBERT WERRER TO EVELVERHAT ERRORESTSYRIOEFFONITS LACLEC
SON SUMVAQAOKDEOVEM LAVI GRELDESPECTUREM INTO FONIETYCT FALTFEROYYVP
DOFLOPILVVNEPDPWPKMRIRKHPILOFSFVOL CHROLOGYAVFALATPORTYCLE TEA FOATHFY SEWELEAA ERWALDAACHDMEETIFMBARCOETEALEE ERRYCH

CPn_0987 1137483 1138115
ytgB-like predicted rRNA methylase
LENGIFAIGFFMFAYRTLLFHINVOVSHBIFKTTVVPGDTVIDATCGNGNDSLFLARLLQ
GEGRLVVYDIQKEALSNALLLFETHLSEQERSVIEMEQSHEHILEKDVKLIHYNLGYLP
KGNKEITTLARTTEISLEYALNIVRPDGLITVVCYPGHPBGEKETHSVESLAQRLHPKEW
CVSSFYVANRCRAPRLFIFQRQGSESSVDKG

MILE AREA TO A CONTROL TO A CON TSAGKLIDAAGLKGLAIGGAQISPLHANFIINTGKATSDEVKQLIAIIQSTLKTQGIDLE HEIRIIPYQPKIHSPVSEK

CPn_0989 1139552 1139016 CTB32 hypothetical protein LRTSLAVKCVLLTIFWLLVMATLSPEKFSGSPISISKEFPQQKMREIILQMLYALDMAPS AEDSLVPLLMSGTAVSQKHVLVALNQTKSILEKSQELDLIIGNALKNKSFDSLDLVEKNV LRLTLFEHFYSPPINKAILIAEAIRLVKKFSYSEACPFIQAILNDIFTDSSLNENSLSI

1139880 1140440

infC-Initiation Factor 3
SVALNFKINRQIRAPKVRLIGSAGEQLGILAIKDALDLAREAGLDLVEVASNSEPPVCKI
MDYGKYRYGLITKKEKDSKKAOHQVRIKEVKLKFNIDENDFSTKLKQARTFVEKCNKVKIT
CMFRGRELAYPEHGFKVVQKMSQGLEDIGFVEAEPKLAGRSLICVVAPGTVKTKKKQEKS HAQDENQ

CPn_0991 1140394 1140612 r135-L35 Ribosomal Protein

KORKNIRKSLMPKMKTNKSVSARFKLTASGQLKRTRPGKRHKLSKKSSQEKRNLSKQPLVD KGQVGMYKRMMLV

CPn_0992 1140622 1140996 r120-L20 Ribosomal Protein GKLVMYAPGSVASRRRKKILKQAKGFWGDRKGHIRQSRSSVMRAMAFNYMHRKDRKGD FRSLWIARLNVASRIHSLSYSRLINGLKCANISLNRKMLSEIAIHNPEGFAEIANQAKKA

CPn_0993 1140975 1142030

"phsS-Phenylalanyl tRNA Synthetase, Alpha"
KSRSHSLGIRISMEMKEEIEAVKQOFHSELDOVNSSQALADLKVRYLGKKGIFRSFSEK
LKGETDKAKLGSLINDFKTYVEDLLQEKSLVLLASEQAEAFSKEKIDSSLÆGDSOPSGGR
HILKSILDDVUDIFVHLGFCVREAPNIESEANNFTLLNFTEDHPARQMHDTFYLNATTVL
RTHTSNVQARELKKQOPPIKVVAPGLCFRNEDISARSHVLFHQVEAFYVDHNVTFSDLTA
ILBAFYHSFFQRKTELRFRHSYFFFVERGIEVDVSCECCGKGCALCKHTGWLEVAGAGMI
HPÖVLRNGNVDPELYSGYAVGMGIERLAMLKYGVSDIRLFSENDLRFLQQFS

HPOVLENGNVDPE1YSGYAVOMGIERLAMLKYGVSDIRLFSENDLRFLQQFS
CPP_O994 1142371 1144440
CTB377 hypothetical protein
LFMERRGGRMKRSRRNFEQALENLEKLKEISLATSNDSYLNNPARFNQRKQTGSSVMEMK
EAŁKIVENYLLEISCVSKSHADKALKESDFLIAGVQNVFSFLENQEDLYKSLLDEYSEVT
KAYBEVKKNLKEVPTYDLSTDEFTEHKEPECFLNNLVEVKRDRSYELFYMLDEQDKFY
NDALVOIIYKONKLHETVNEGDPLTKTLLWNSEEVKNIASSLVIVNDMPLRLFYQRALSH
LDIEAVVKYHNAVMALFFSRYEATMVFKSPKKNNIWYFNDFLLFLREAMKDLNNVLDSQ LDJEAVVKVHNAVMALFFSRYEATMYFKSPKKHNIWYFNDFLLFLREAMKDLINNNYIDSQ
ERROTKLLASALSLGIFESKLVFEEASRYLYFNIQTKLENANGKKPLSPGQYLTDAYEEL
HRLESKYPNGPLFKAMDRVLEHESRPYDPMILGILPSLEGTLKLHGKSIDIIRSPSPVTQ
SSĪLYANCNEEFLGFLNAKAHRSEVTLVLNIQNRISRKERARSRVIEEALEQEEHAPYVH
AFŠEPEPEELLONLESIHGDIETFADFFSILQEEFHKPLLASSFFLTKELKEFVGSFLKE
KLŢĀLKDIFFAKKKILFRNDKLLLLHLLSYLIVFKLIERTNPNSIVVVSKDGLDYVSVFI
AGFÆFFSREAFWDEHSLKLLLINVLSPTLVARDRLVFVSHIELLSKFVNCLKKNRQGFSS
LKŠFFKDDIEGWEFTGYLHELTEVSHKHNL

CPN_m0995 1145515 1144415
CT038 hypothetical protein
RMLEWKRHELTREWFALTSLLVLALIFYASIHHSLHTLKGASTAASGASVKLSILYYLAQ
ISĒKĀEFLMPQDLVAVATTSTLFAMQNKRBIILLQASGLSLKSLMHPLLLSGAVIMMVLYA
NFONLHPICEKISITKENMDRGTTDKEÇGKIPALYLKDQTVLLYSSIEPKTLTLNNVFWI
KDPKTIYTMEKLAFTTLSLPIGLNVTOFFANDSENLELKEFFDMKEFPEIEFNFYENPFS
KLFSAGNNNRLSEFFRAIPWNATGLGLSTQVPQRILSLLAQFYYVLISFLACMAAIILSA
YLCLRFSRTPTVTLAYLIPLGTVNIFFVFLKAGIVLASSSVLPTLPVMAFPLIVLFLLTN
VAVAUKO

YAYAKLO

CPn_0996 1146592 1145519
CT839 hypothetical protein
AMPILWKVLIFRYLKTAAFCTLSLICISIISSLQEIVAYIAKDVPYDTVLRLMAYQIPYL
LPFILPGSCFVSAFSLFRKLSDNNHMTFLRASGASQSIIMFPVLMVSGAICCLMFYTCSE
LASIGRYQTCKEIANMAMTSPALLLQTLQKKENNRIFIAVDHCAKSKFDNVIVALKGNNE
ISHVGIIKŠIIPDTTKDTVAKDVVFISKLPDSLTESSSPSSQRFYIETLDELLIPKITS
TLFAGKSVLKTRTDYLPWKQLVKQSLKHSHLPETLRRVAIGFLCITLTYAGMILGIHKPR
FRKSIALYFIFPILDLILLIVGKNTKNLPLAFMLFVFPQLVSWVVFAARAYRESRGYA

CPn_0997 1146699 1147664
mesJ-PP-loop supertamily ATPase
AYKMYLSSOLLRODKQLDLFFASLOVKRRYLLALSGGSDSLFLFYLLKERGVSFTAVHID
HGWRSTSAQEAKELEELCAREGVPFVLYTLTAEEQGDKDLENQARKKRYAFLYESYRQLD
AGGIFLAHHANDQAETVLKRLLESAHLTNLKAMAERSTVEDVULLRPLLHIFKSSLKEAL
DARGISYLQDP::MEDERYLRAMRKKLFFWLEEVFGFMITTPPLLTLGEESAELSEYLEKQ
AQFFGAATHQD::QGELPCPDCLIQQAFLCKWVMKKFFNNAGIAVURHFLQMVYDHLSRS
CCATLEMRNKIVIIKTGVVVID

CEN_0998 II478HI II50584

TOSH ATP-dependent zinc protesse
LIGHVKEMSKOKKMKEEPKKNEETVEMFELLFYJVFFYZYAPQNFLAGKKARVOFSHQIEH
LYNLKLIVPEDSHKIALNDHLVSFGGRERDVOTQEGGLPYHYLELLDXSHRLDLDLQETG
KSHTHIAKEVTNSHLMFSA 1835BIPEGGYA ISYPSEZGSWISTDELVVTGAATPQLINL
GEQQEPYTLGRSPFALRTYGDDLYELLIGKYLSTVIA FOSTLKRELKDLYQOVEVSTVI
ETTZFAAYTLYJQVISTLNR ISSSLVVSEGGERFSQLFGWKLYREIMNYHKLVFARDLN
QAQLEKLRGELSOTVWYFNNQELSSRSLEKODPEVFYSWFAGAKEIWTAFKFNHILSFKA

PDQPRNLVLEKTFKSGEPSPHYLGYLFTFLFIILVLLFVYLVFSRQMRGMSGSAMSGKS PARMLLKGONKVTFADVAGIEEAKEELIEIVDFLKNFNKFTSLGGRIPKGVLLIGPFGTG KTLIAKAVSGEADRPFFSIAGSDFVEMFVGVGASRIRDMFEQAKRNAPCIIFIDEIDAVG RHRGAG ICOGHDEREOTLNCLLVEMDGFGTNEGVILMAATNRPDVLDKALLRPGRFDRRV VMNLPDIKGRFEILMVHAKRIKLDPTVDLMAVARSTPGASGADLENLLNEAALLAARKDR VMNLPDIKGREEILMYHAKIKLDEYULMAVARITKASSADLENLLNEAALLAARNOK TAVTAVOVAEADKVLYGKERRSLEMDAEERKTTAYHESGAHAVUGLUUQHGDEYUKVTII PRGLSLGATHFLPEKNKLSYWKKELYDQLAVLMOGRAAEEIFLGDISSGAQQDISQATKL VRSMYCEWIMSEOLONYTYDERSDGLTGYGGYHEKSYSEETAKTIDTELRMLLDAAYQRA 18 AN IEBUMTUML TO: FONKEIMFUTACOLL VARU DE IMPERKIJSEEL

CPn_0999 1152859 1150766
pnp-Polyribonucleotide Nucleotidyltransferase
QETFMNFQTISINLTEGKILVFETGKIARQANGAVLVRSGETCVFASACAVDLDDKVDFL
plrvbyQEKFSTGKTLGGFIKREGRPSEKEILVSRLIDRSLRFSFPYRLMQDVQVLSYV
SYDGQVLPDPLAICAASAALAISDIPQSNIVAGVRIGGIDNQWVINFYKTELASSTLDL
VLAGTENAILMIEGHCDFFTEEQVLDAIEFGHKHIVTICKRLQLMQEEVGKSKNLSAVYP
LPAEVLTAVKECAQDKFTELFNIKDKKVHAATAHEIEENILEKLQREDDDLFSSFNIKAA
CKTLKSDTMRALIRDREIRADGRSLTTVRPITIETSYLPRTHGSCLFTRGETQTLAVCTL
GSEAMAQRYEDLMGEGLSKFYLQVFFPPFSVGEVGRIGSFGRREIGHGKLAEKALSHALP
DSATFPYTIRIESNITESNGSSSMASVCGGCLALMDAGVPISSPIAGIAMGLILDDQGAI
ILSDISGLEDHLGDMDFKIAGSGKGITAFQMDIKVEGITPAIMKKALSQAKQGCNDILMI
MNEALSAPKADLSQYAPRIETMQIKPTKIASVIGPGGKQIRQIIEETGVQIDVNDLGVVS
ISASSASAINKAKEIIEGLVGEVEVGKTYRGRVTSVVAFGAFVEVLPGKEGLCHISECSR
QRIENISDVVKEGDIIDVKLLSINEKGQLKLSHKATLE QRIENISDVVKEGDIIDVKLLSINEKGQLKLSHKATLE

CPn_1000 1153193 1152891
rs15-S15 Ribosomal Protein
SAFAAIILRRHPMSLDKGTKEEITKKFQLHEKDTGSADVQIAILTEHIAELKEHLKRSPK DONSRLALLKLVGQRRKLLEYLNSTDTERYKNLITRLNLRK

1153369 1153869

yylelggeklinmekdiffmqqafkearkaydqdevpvgcvivkddkiiarahnsveklk Datahaeilcigsaaqdldnwrlldtvlyctlepclmcagaiqlariprivwaapdvrlg AGGSWVNIFTEEHPFHTVSCTGGVCSEEAEHLMKKFFVEKRREKSEK

CPn_1002 1153844 1154089 CT845 hypothetical protein KSAERKVKNKIVTLLDQLYEDQESRLQKLGEEIVPNLTPEDLLQPMDFPQLEGNPAFRFE EGVLSGIGEVRAAILAALSQEN

1154862 1154092 CPn_1003

CPT_1003

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T154052

T1 LDFFLFYKAIOHLIRKLGAFSVLVVISGOALIIGAVLWGFMALIHSSQSFFGPESIICGV LTVOIFLDPEKRFTIGPTPLSVSIKWGFLFVLGFYCCILIFSGAFLLLLASMLAIVLAIL FCKKEKIPNPYTTSLRF

CPn_1004 1155418 1154879 CT847 hypothetical protein HLSIEELMSIQPVSNTTTKADKVIPDSTKVISDSITINKQSAFYFCISVMLRLSESTTEY GKSILAVLEDNTIVQQQRVKELINLPLLKVPDLQKKDGSDDEYKNQNEIQAYQSSNQQIS ANRQMIQQELSSAQQRAQANQKSVNSTTIESMQILQATSSMLSTLKELTIKANLTNSPSD

CPn_1005 1155957 1155415
CT848 hypothetical protein
nRKFVRLMMUIDPLSAKKPLQAAINVPGTPITGGPNTATADDIIAKFSKDSNPLIVTVY
YVYQSVLVAQDNLSIIAQELQANSSAOTYLNNQEALYQYVSIPKNKLNDNSSSYLQNIQS
DNQAIGASRQAIQNQISSLGNAAQVISSNLNTNNNIIQQSLQVGQALIQTFSQIVSLIAN

1156493 1155990 CPn_1006

CT849 hypothetical protein
TKVNFFIMSITTLGTLPTVNTINSSRPPLEPLNTPKIGAVLFSIYELLLQAIEIRQQTVL tosoolndntniqoolnoetnoikyaivsagakedeitrvonononysaorsniqdelvt trongoiilshastniniiqoossodssfikttnsigstvnolnkplg

CPn_1007 1156689 1156907 CT549.1 hypothetical protein LWYKSLAGEEKDVSGNECNDYPEVFKDDVSAYVLVTCGQMSSEGKIQVEMTYEGDPAVIS YLLTKARDSLDES

1156904 1158223

CT84850 hypothetical protein
VLNYSFIGMLKPMYVLSKRLYRWWNQLIKLGDLVKNSRSFSVEWVFISALLLIFGCLGCA
SVVKVSLVPFLLLFSFLAFFLILCFRGKGYALLLGVFVTLYVAKYVVGETLYVSFVLGGL
GVSFLLAFGLFLGGWMLAQEEEMVKGKEQLRLSEDLDAQRSAYEDLLLTKSQEKEFLDAR
AGGLDPELTECQELLKAAYCKQEYLTIDLKILADQKNSWLEDYAELHNKYIELVSKNGDV VFPWVAEPSVOESQOSERVDVSRWVSALQEKEESLERLRNEILVEKQRCSDYEHRCOELG LLLQNFTALERRCEELQNLLNQKETQINELHOLVCKSEEKVSVEPSAHAETSCVEEKQYK GLYSQLQEQFLEKSETLSLVRKKLFAVQEKYLTLKKKEELTKQDISFDDISMIQGLLERI EILEEFYSHLEELVSRSLSL

CPn_1009 1159085 1158186

map_Methionine Aminopeptidase
YRLCHEVILMKRNDPCNCGSGRKWKQCHYPQPPKMSPEALKQHYASQYNILLKTPEQKAK
IYNACQITARILDELCKASQKGVTTNELDELSGGELHKKYDAIAAPFHYGSPPFPKTICTS
LHEVICHGIPNDIPLKGDINNIDOWGIVDGYYGDCSPMYMIGEVPEIKKKICQAALECL
NDGIAILKPGIPIXELGEAIEARADTYGFGVVDQFVGHYMGIEPHENPYVPHYRNSMIP
LAKAMIFTEEPMINVGKKEGVVDPKNQWEARTCDNGPGAQWEHFIAITETGYEILTLLND

1159675 1159067

PERGANDERER LATERITOR P. 112-2009. TERROADERER LATERITOR PROPERTY OF CONTROL OT CONTROL OF CONTROL

/ In_1011 Hed309 1153902

1159402

CTHS: hypothetical protein EMRLKNYPMIQESFFLPQTCILLLASDSLTNILALHHULANYSVKQRMLVLLRESFFAFI AMFALYGLALXGLKVLNTPVCAIEVVGGIAVTLAGVRAVLRLJKEESWIPYKFNMSPSYS PCIGPIALPLMFGPSG

CPn_1012 1162220 1160421
yzeB-ABC transporter permease
ALFGLITSKMKKKFIFYEYTVEGLLELWEMTSRHPPTEGFFCPPEGGLAGGTLOSLPLLL
TOWNHERANDERG TOWN TOWN TOWN A PROPERTY OF A MAKE OF THE
POWER DAVE PROPERTY TO THE CONTEMPERTY OF THE PROPERTY OF THE PRO HIFSGLKIAIGSAGFAAIAGEWVASQSGLGILMLESRRNYEMELAFAGLATLSILTLSLF QITLLIEKLIFSLFRVRRMSLKHKSVAKKALSVLALIPIMLIPWKGNSKSPPDKKNLTSL TLLLDWTPNPNHIPLYAGVAKQYFKQHGLDLQLQKNTDSSSAVPHVLFQVDMALYHALG IMKTSIKGMPIQIVGRLIDSSLQGFLYRSQDFTYKFEDLNGKVLGFCLNNSRDLNRLLET

LNRNGVVPSEVRNVSSDLISFMLLNKIDFLYGAFYNIEGVKLOTLOMPVKCFLSDTCDLP TGPOLIVFTKKGTKASEPEIVEAFOKALOESIIFSKDHPEDAFKLYAKETKSIPKNLYQE YLQWEETFPLLAQSQDPLSKDLVDKLLETIIKRYPELASEVAKFSLNDLYNPSLPEEGSV

1162209 1163624

CPn_1013 116209 1163624
fumC-Fumarate Hydratase
RENSINHRONIDMRQEKDSIGIVEVPEDKLYGAQTMRSRNFFSWGPELMPYEVIRALVWI
KKCAAQANQDLGFLDSKHCDMIVAAADEILLEGGFEEHFPLKVWGTGSGTGSNMNVMEVIA
NLAIRHHGGVLGSKDE JHPNDHVNKSQSSNDVFPTAMHIAAVISLKNKLI PALDHHIRVL
DAKVEEFRHDVKIGRTHLMDAVPMTLGQEFSGYSSQLRHCLESIAFSLAHLYELAIGATA
VGTGLNVPEGFYEKI HYJLRKETDEPFIPASNYFSALSCHDALVDAHGSLATLACALITKI
ATDLSFLGSGPRGGLGELFFPENEPGSSIMPGKNPPTOC EALQMVCAQVUGNNCTVIIGG
SRGNFELNVMKPVIIYNFLQSVDLLSEGMRAFSEFFVKGLKVNKARLQDNINNSLMLVTA
LAPVLGYDKCSKAALKAFHESISLKEACLALGYLSEKEFDRLVVPENMVGNH

1165456 1163732

ychm-Sulfate Transporter ALASTLGYCIVKVPWAFKNFIPKLYTSIKEGYSFNTFKKDFQAGITVGILAFPFAIAIAI ALASTLGYCIVKVPWAFKNFIPKLYTSIKEGYSFNTFKKDFQAGITVGILAFPFAIAIAI
GVGVSPIQGLLASIIGGLLASAMGGSNVLISGPSSAFISILYCLSAKYGAEALFTVTLLA
GVFLIAFGLTGLGTFIKYMPYPVYTGLTTGLAIIIFSSQIKDFLGLQMGANIPADFLPKW
IAYWDHLWTWDSKSFAVGLFTLLIMIYFRNYKERYPGYMIAIVTATTLWWLLKIDIFTIG
SRYGTLPTAIPLEKIPQLSITKLIQLMPDALTIAVLSGLETLLSAVVADGMTGMRHGSNC
QLVAQGVANIGTSLFSGIPVTGSLSRTAASIKSEATTPIAGIVHSIFICFILLLAPLTV
KIPLTCLAAVLLIJAWMMSEIHHFIHLFTAPKKDIVVLLTVFILTVMTTITAAVQVGMML
AAFLFMKOMSDLSDVISTAKYFDKDSDFLSKAEVPONTEIVEINGPFFGIADRLKNILN
DIEKPPKIFILCMTRVPTIDASAMHALEEFFLECDRQGTLLLLAGVKKTPLADLKRYHLD
ELIGVDHIFSNIKSALLFAQALTNLESKTSTRHLV

CENTIO15 1165550 1166893
CT857 hypothetical protein (possible IM protein)
KNYKNFSFFTSVRVRSKVDHEIILEVTMLKLOLCALFLFGYLAIVFEHIVRVNKSAIAL
AMGSLMWLVCFSHIPMADHMILVEEIADMSQVIFFLFSAMAIVELIDAHKGFSVIVKFCR
IQSRPILLWALIGLSFFLSAALDNITSIIIIISIIKRLIVKAREDRLLLGAICVIAVVANGG
AWTPLGDVTTTMLWINNKITSWGIIRALFVPSLVCVLVAGFCGQFFLRKRGSTLIAKDVE
LQSAPPKSLWIIFIGLGSLLMVPWKACLGLPPFMGALLGIGLWWLTSDWIHSPHGEDRY
HLRVPHILTKIDISSITFFIGILLAVNALSFANLITDFSLWMDKIFSRNVAIVIGLLSS
VLDNVPLVAATMGMYTLPLDDTLWKLIAYAAGTGGSILIIGSAAGVAFMGLEKVDFLWYF
KRISWIALASYFGGLFSYFVLESLNFFI

CPTL1016 1167027 1168898
CTREAS hypothetical protein
KREVPMKKGKLGALVFGLLEFTSSVAGFSKDLTKDNAYODLNVIEHLISLKYAPLPWKELL
KREVPMKKGKLGALVFGLLEFTSSVAGFSKDLTKDNAYODLNVIEHLISLKYAPLPWKELL
KSEDGHVFVVDVOTSQGDIYLGDEILEVDGMGIREAIESLBFGRGSATDYSAAVRSLTSR
SAARGDAVPSGIAMLKLRRPSGLIRSTPVRWRTTPEHIGDPSLVAPLIPEHKPOLPTOSC
VLFRSGVNSQSSSSSLFSSYMYPYEWELRVONKORFDSNHHIGSRNGLPFFGFIJHWEQ
DKGPYRSYIFKAKDSQGNPHRIGFLRISSYVWTDLEGLEEDHKDSPWELFGEIIDHLEKE
TDÄLIIDOTHNPGGSVFYLYSLLSMLTDHFLDTPKHRMIFTDDEVSSALHWODLLEDVFT
DEDANAVLGETMEGYCMDMHAVASLQNFSGSVLSSWYSGDINLSKYMPLLGFAQVRPHPK
HQYTKPLFMLIDEDDFSCGDLAPAILKDNGRATLIGKPTAGAGGFVFQVTFPNRSGIKGL
SLTSSLAVRKDGEFIEDLGVAPHIDLGFTSRDLQTSRFTDYVEAVKTIVLTSLSENAKKS
EECTSPOETPEVIRVSYPTTTSAS EEQTSPQETPEVIRVSYPTTTSAS

EEGITSPOSTPEVIRVSYPITISAS

\$\frac{1}{2}\limins\$

CPn\[\frac{1}{2}\limins\$

CPn\[\frac{1}{2}\limins\$

1168997 1169935

1y\frac{1}{2}\limins\$

VIIMRKLILONPRGFCSGVVRAIQVVEVALEKWGAPIYVKHEIVHNRHVVNALRAKGAIF

VEELV\[\frac{1}{2}\limins\$

VEELV\[\frac{1}{2}\limins\$

VEELV\[\frac{1}{2}\limins\$

VEELV\[\frac{1}{2}\limins\$

VEELV\[\frac{1}{2}\limins\$

VEELV\[\frac{1}{2}\limins\$

VEELV\[\frac{1}{2}\limins\$

VEELV\[\frac{1}{2}\limins\$

VEELV\[\frac{1}{2}\limins\$

VEELV\[\frac{1}{2}\limins\$

LIGHKKHVEVIGIVGEVPEHITVVEKVADVEALPFSSDTPLFYITQTTLSLDDVQEISS

ALLKRYPSIITLPSSSICYATTNRQKALRSVLSRVIYVYVVGDVNSSNSNRLREVALRRG

VPADLINNPEDIDTNIVNHSGDIAMTAGASTPEDVVQACIRKLSSLIPGLQVENDIFAVE

DVVFQLPKELRCS

CPn_1018 1169895 1179629 No robust homolog present in Genecank/EMBL as of 11/7/98 RMSYFNYQKNSVVLRCLGLLAKFFSRLLYPVFF5FREGIYLFSSLYLKYPRLFFYDLGKY VYSLRHCPYAKLGRLPGASLLKEGNYYGETPWSVLAFICQAFDITSGDILYDLGGGIGKV CFWFSHVVRCQVIGIDNOPHFIRFSSNMHRKLSSGFALFDTEEFKNVVLSQASYVYFYGS SFSRRLLNEIILKLSEMAPGSVVISISFPLDSFSRGKECFFTEKSCSVRFPWGKTIAYKN

CPn_1019 1172146 1170638

CT860 hypothetical protein
IHRRNIMTVSYQSISTPPPEGEFTIFVDONATEEAV/AAEVQVALPAGEQYAMLRATSEL
CFG LLTGGECALFOAL PPKERPLQEEQFLVMOGILMEGTSLPNLKEGQSQCTSLASHRNP
LAQQTSUSMSTNKASTETTSSSFPFFSCK\PEGDSC/DKTFTVSVQTPKAQEQQEASASQ
SQAQFHVRSYGUSTIKHSHAKEKYOSOSTKGALFOKHTOTKSDATLSPMSLYCTLHKEVPQ
ALSCREGOKKEEHROLGQGEGYEGEGREGEGKKTFWTUSLCQYTSJSNQVYESYTPI
IPDF FVEFALDESQLJVLAGRVTNLDVLPICTELMYLMIKGRANDFMTRLEEPELMERE
AHBLAGJSROAKYARWIGITATATLGILGAIAPMVGEIGGDILGFVQRISGPFKDATAK
TEFKGIJEVPTSLGQLGEAASKVHELJET\VRAV\FZPKEVPPMPQDEVTRTIEEVKDNW
KOMUNELINILQTEHDAARSLEG

GRATMOEDAMERGEBALGERYTDRIFFGORDES GREENLETE OUR ELANDOETA ZORF EGOOGREEHELLIERKEL ELYLALEER ET BEGESTE OOSEEL OF HELD OEK DANDE OOR OEK DE SKRALDA OELANDE SOR OF BREELD OEK DE STANDE OOR OEK DE SKRALDA OELANDE SOR OEK DE SKRALDA OELANDE SOR OEK DE SKRALDA OELANDE SOR OEK DE SKRALDA OELANDE SOR OEK DE SKRALDA OELANDE SOR OEK DE SKRALDA OELANDE SOR OEK DE SKRALDA OELANDE SKRALDA OE

DSEQQELLQSRREERSETYANQQSSEKKIETKVQIKDLCKDLFSQDQDSNGKGKKKSFPQQ DTGRKNRIAKAAQAVPVIPPPSIGVFTLSYLLTKQGILSDFSSYCCHKDSVESTQRELDA LHERRIETIKVSIEKEKRERLWGSLSDIIGWLAPFVSIGIGIVAILSGGGIFAFAGFFAG LISLVIKCLEKLKFWWLEKHLPIKNEELRRKIITIIGWVYVLTPVILSICTLKVEPLGF SPIIEGAIKGIQPAIESTMAALRCAILFSQAEIYKLKGKLTKIQLDIELKSFDRDDHYER SQELLDNMESSFEALSRILNYMRELDQVYLHSLRG

1174270 1173698 CPn_1021

The second of th TYQKIFKISSEDLEKYYKEGYHAYLDKDYAKJITVFRWLVFFNPFVSKFWFSLGASLHMS EQYSQALHAYGVTAVLRDKDPPPHYYAYICYTLTNEHEEAEKALEMAWVRAQHKPLYNEL KEEILDIRKHK

CPn_1022 1175709 1174216 CT863 hypothetical protein FSFFFYALKLQIMNMPVPSAVPSANITLKEDSSTVSTASGILKTATGEVLVSCTALEGSS STDALISLAGGILATQELLLQSTNYHQLLFLPEVVELEIQVVDLLVQLEHAETITS
EPQETQTGSREEQTLPQQSSSKQSALSPRSLKPEISDSKQQQALQTPKDSAVRKHSEAPS
PETQARASLSQASSSCRSLPPQESAPERTLLEQQKASSFSPLSQFSAEKQKEALTTSKS
HELYKERDQDRQQREQHDRKHDQEEDAESKKKKKKRGLGVEAVAEEPGENLDIAALIFSD OMRPPAEETSKKETTFKKKLPSPMSVFSRFIPSKNPLSVGSSIHGPIOTPKVENVFLRFM KLMARILGQAEAEANELYMRVKQRTDDVDTLTVLISKINNEKKDIDWSENEEMKALLNRA KEIGVTIDKEKYTWTEEEKRLLKENVQMRKENMEKITQMERTDMQRHLQEISQCHQARSN VLKLLKELMDTFIYNLRP

CPn_1023 1176008 1176331 No robust homolog present in Genebank/EMBL as of 11/7/98 GLDFLEIFIMKKVVTLSIIFFATYCASELSAVTVVAVPLSEAPGKIQVRPVVGLQFQEEQ GSVPYSFYYPYDYGYYYPETYGYTKNTGQESRECYTRFEDGTIFYECD

1177317 1176334

xerD-Integrase/recombinase IFFFPWFSLCSLKIAPLPILKLHSLASMTMPSTQFHTTILEQFSLFLSVDRGLCQQSIAA YRODISSPLTISAISSPODISONSVYIFAEELYRRKEAETTLARKLIALKVFFLFLKDQ LLPYPPITEHPRIWKRLPSVLTPQEVDALLAVPLQMEKNPRHLAFRDTAILHTLYSTSVR VSELCDLRIGHVSDOCIRVTGKGSKYRLVPLGSRAFRAIDAYLCPFRDQYOKKNPHEDHL FLSTRGHKLERSCVWRRIHNYAKQVTSKPVSPHSLRHAFATHLLDNKADLRVIQEMLGHA RIASTEVYTHVAADSLIEKFLAHHPRNL

1177266 1178879

CPH_1025

gg1-Glucose-6-P Isomerase

GAEOFSSYREKTMERKRFIDCDSTKILQELALNPLDLTAPGVLSAERIKKFSLLGGFTF

SFATERLDDAILAALISLAEERGLHESMLAMQQQQVVNYIEGFPSEMRPALHTATRAWVT

DSSFTGEAEDIAVRSRVEAQRLKDFLTKVRSQFTTIVQIGIGGSELGPKALYRALRAYCP

TDKHVHFISNIDPDNSAEVLDTIDCAKALVVVVSKSGTTIETAVNEAFFADYFAKKGLSF

KDHFIAVTCEGSPMDDTGKYLEVFHLWESIGGRFSSTSMVGGVVLGFAYGFEVFLQLLQG ASAMDQ IALQPNARENLPMLSALISIWNRNFLGYPTEAVIPYSSGLEFFPAHLQQCCMES NGKSIAQDGRRVGFSTSPVIWGEPGTNGQHSFFQCLHQGTDIIPVEFIGFEKSQKGEDIS FQGTTSSQKLFANMIAQAIALACGSENTNPNKNFDGNRPSSVLVSSQLNPYSLGELLSYY

ENKIVFQGLCWGINSFDQEGVSLGKALANRVLELLEGADASNFPEAASLLTLFNIKFR CPn_1026 1178961 1179137

CSFGFGKICEDRMFFIAVRSRGFLDIHGILAARKGKQVVKSTAGAWIGSRGAVFYSLVS

CPn_1027 1179172 1180755
No robust homolog present in Genebank/EMBL as of 11/7/98
NMPGSVSSPPLSPVIVRERVPSSSGSDLIQPHAVLKISILIFALVTILGIVLVVLSSALG MMGSVSSPPLSPVIVRERVPSSSGSDLIQPHAVLKISILIFALVTILGIVLVVLSSALG
ALPSLVLTVSGCIAIAVGLIGLGILVTRLILSTIRKVDAMGYDAAVKEEQYLSRIRELES
ENREIRDRNRAVEDCCAHLSEENKDLRDEYYLHGMTERLIASLEIENQALVAENILLKDW
MASLSRDFRAYKQKFPLGALEPWKED IACIMEQNLFLKPECIAMVKSLPLETQRLFLYPK
GFQSLVNRFAPRSRFFQTBKYEYNSRNENEDGKVAAVCARLKKEFFSAVLGACSYEELGG
ICERAVALKETLPLPEAVYDTLVQEFPNLLTAESLWKEWCFYSYPYLRPYLSVDYCKRLF
VQLFEELCLKLFTTGSPEDQALVRLFSYYRNHIPAVLASFGLPPPETGGSVFVLLPKQEN
LLWSQIEVLATRYLKDTFVRNSEWTGSFEMMFSYNEWCKELSEGRIRFAEDYETRHSEEF
BDSSIESPEGETEFF DECEMERYGIR ENDRY UNDSWAMBURDWEDG PPSPLSEEGEGEFLPPCSEEEVSVLERPDLDVDSMWVWHPPVPKGPL

1180995 1181999

CPT_1028 1180995 1181999
mdhC-Malate Dehyrogenase
FFLKGVRMAFKEVVRVAVTGGKGQIAYNFLFALAHGDVFGVDRGVDLRIYDVPGTERALS
GVRMELDDGAYPLLHRLRYTTSLNDAFFGIDAAFLIGAVPRGPGMERGDLLKGNGQIFSL
GGAALNTAAKRDAKIFVVGNPVNTNCWIAMKHAPRLHRKNFHAMLRLDQNRMHSMLAHRA
EVPLEEVSRVVIWGNHSAKQVPDFTQARISGKPAAEVIGDRDWLENILVHSVQNRGSAVI
EARGKSSAASASRALAEARSIFCPKSDEWFSSGVCSDHNPYGIPEDLIFGFPCRMLPSG

DYEIIPGLPWEPFIRNKIQISLDEIAQEKASVSSL

CPn_1029 1181987 1182844

No robust homolog present in Genebank/EMBL as of 11.7798

RVFVISTMLWGYSMRQSFDELSQNAFKNIFNKQPFCFIFCSLCCFGFVFALFLKLCSRLA

PEISLSTLGLGAFFCAFSVICASAIIVQFLLHKESQGETSKLCCAIKNTWSSLWLSLLVS

MPFFIAMVAVVTVAMLSSFLGSLDWVGKLFHTVLIFIPYLSATALILLFLGSFSCLFFCI

PVLHNQESIDYRKLLECFRGNILRQFIGVVIALVPLALCSWLALDSFYLMTHLVEIADIH

TWSFLAQMFVLIVPIALILTPAVSFFFNFSFSFYLAKQEEEKALVK

CPD_1030 1183901 1182843
predicted D-amino acid dehyrogenase
FKVHFMRIAVUGAGYAGUSVTWHLLLHSGGTATIDLFDPIPLGEGASGMSSGLLHAFTGK
KALKPPLADQGINATHALITEASKALNVPIVISGGIURPAIDEDQAQLFTERVEEFFKEV
EWWEKAPCEISIPSMVIPPNUGALPIKCOVTLMIDLYIQCDADACMKLYTQFYDELIEDL
ADIEEFYDHITYTDGANASTLPELKDMFVNKVYGQLLEICMPFDLAMLSFSINAHKYMVA
NTQKNTYIMATFEHNQPECTPDPAIAYQEIMFPYDGLFFGLKDAQVLHCYAMRSSKS
RLPVIJPIPEKLWCLAGUGSKGLLYHGITGDMLAQAVURFOTAYIAKEPUFTI

1185567 1184098

CFU_10:1 TERSON TERRORS

ARECO ARGUILLER AGELEPHER AGELEPHERE
ERFTEMPERTKJORNEGT LALAGMVVOOLEGEREPELEPHERATAGAGAVILJWILTTOPG
MET LAFTER LUTELREDUKRGI EMIZO ERFOLEPHERWYMERO LEGALVA LEFT LE
ENTET PERFOCKINTED ALIANS TELIMORIE TO VIZO LEGALVE LEFT LE
TAFFER LAVERTOPWIRTEN TRADEROUTONG LOVITHUNGT LOVITH LOT LEGALVE MORRAND
LINGONTYLOFERS LEFT LEGALGEROUTONG LAGER PERFOTAGVED LLAGRWIDEVLINN

OLITAVLSOWLGWTIIVAEIPFSAAKNOTFPEIFTIENKEKSPSVSLYITSSVMQLAMLL VYFSGNAWNTMLSITSVMVLPAYLASAAFLFKLSKSKTYPKKGSIKAPLAMITGILGVVY SLWLIYACGLKYLFMALVULALGIPFYIDAGKKKKNAKTFFAKKEIVGMTFIGLLALTAI

1186153 1185566

CT373 hypothetical protein
[MAYSTPYPTLAFHTGGGEGDDGMPPGPFETFGYDGALLQAKIENFNIVPYTSVLPKEL RETURNS METERINE TO ANALYS TO MAKE ENVES DE METERINE METERINE DE MAKE PER PRODUCTION DE LA CONTRACTION DEL CONTRACTION DE LA CONTRACTION DE LA CONTRACTION DE LA CONTRACTION DE LA CONTRACTION DE LA CONTRACTION DE LA CONTRACTION DE LA CONTRACTION DE LA CONTRACTION DE LA CONTRACTION DE LA CONTRACTION DE LA CONTRACTION DE LA CONTRACTION DE LA CONTRACTION DE LA CONTRACTION DE LA CONTRACTION D ALGFLNFENAEPAKVN

CPn_1033 1187656 1186187
CT372 hypothetical protein
NNKKDYSGEFLITDTUDSIAFLPSEENFCYIKTILFFRVKKKHYAFFYGEFMISFRFLL
LSGLCALGISSYAETPKETTGHYHRYKARIQKKHPESIKESAPSETPHHNSLLSFVTNIF
CSHPWKDGISVSNLLITSVEKATNTOISLDFSILPOWFYFHKALGGTQALEIPSWOFYFSP
STTWTLYDSPTAGOGIVDFSYTLIHYWQTNGVDANQAAGTASSMDYSNRENNLAQLTFS
QTFPGDFLTLAIGQYSLYAIDGTLYNDNOQYSGFISYALSQNASATYSLGSTGAYLQFFPN
SEIKVQLGFQDSYNIDGTNFSIYNLTKSKYNFYGYASWTPKPSCGDGQYSVLLYSTRKVP
EQNSQVTGWSLNAAQHIHEKLYLFGRINGATGTALPINRSYVLGLVSENPLNRHSQDLLG
IGFATNKVNAKAISNVNKLRRYESVMEAFATIGFGPYISLTPDFQLYIHPALRPERRTSQ
VYGLRANLSL VYGLRANLSL

CPn_1034 1188589 1187732
Predicted OMP (CT371) [leader (18) peptide]
KTSWOKYKKYLSYSILVOKIARYYMKTWLFFTFLFSCSSFYASCRYAEVRSIHEVAGDIL
YDEENFWLILDLDDTLLQGGEALSHSIWKSKAIQGLQKQGTPEQEAWEAVVPFWIEIQEM
GTVQPIESAIFLLIEKIQKQSKTFVYTERPKTAKDLTLKQLHMLNVSLEDTAPQPQAPL
PKNLLYTSGILFSGDYHKGPSLDLFLEICTPLPAKIYIDNOKENVLRIGDLCQKYGIAY

FGITYKAQELHPPIYFDNIAQVQYNYSKKLLSNEAAALLLRHQMHE

CPn_1035 1190081 1188570
aroE-Shikimate 5-Dehyrogenase
VVQLPLMVPIVHLQIWRFSMIYYGVSVMLCATVSGPSFCEAKQQILKSLHLVDIIELRLD
LINELDDQELHTLITTAQNPILIFFROHKEMSTALWIQKLYSLAKLEPKWMDIDVSLPKTA
LOTIRKSHPKIKLILSYHTDKNEDLDAINNEHLATPAEIYNIVLSPENSSEALNYIKKAR
LLPKPSTVLCMGTHGLPSRVLSPLISNAMNYAAGISAPQVAPGQPKLEELLSYNYSKLSE
KSHIYGLIGDPVDRSISHLSHNFLLSKLSLNATYIKFPVTIGEVVTFFSAIRDLPFSGLS
VTMPLKTAIFDHVDALDASAQLCESINTLVFRNQKILGYNTDGESVAKLLKQKNISVNNK
HIALVGAGGAAKAIAATLAMGGANLHIFNRTLSSAALATCCKGKAYPLGSLENFKTIDI
INGLPPEVTFPWRFPPPIVMDINTKPHPSPYLERAQKHGSLIIHGYEMFIEQALLQFALW
FPDELITPESCDSFRNYVKNFMAKV

1191180 1189984

CPHELU36
argb-Dehyroquinate Synthase
GYDRPCSCRSCIIPTMLQTMMSETIITTPHVVKLISNFFQKKLFSSISTAYPLVIITDVS
VQGHLLGPILDHIKMLGYQVIVLTFPPGEPNKTWETFISLQYQLVDQNISPPKSSIIGIGG
GTYLEMTGFLAATYGRGLPLYLIPTTITAMVDTSIGGKRGINLRGIKNRLGTFYLPKEVW
MCPQFLSTLPREEWYHGIAEAIKHGFIADAYLWEFLNSHSKMLFSSSQILHEFIKRNCQI
KARLVAEDPYDRSLRKILNFGHSIAHAIETLAKGTVNHGQAVSVGMMIETRISLAEGVMK
TPQLIDQLERLLKRFNLPSTLKDLQSIVPEHLHNSLYSPENIIYTLGYDKKNLSQHELKM
IMTEHLGRAAPFNGTYCASPNMEILYDILWSECHVMRHC

1192286 1191123

CPHE 037 1192286 1191123
aroc-Chorismate Synthase
LHFSRGSRRSFLEELLRTSVSRSHYLVKVMKNSFGSLFSFTTWGESHGPSIGVVIDGCPA
GLELHESDFVPAMKRRRRGNEGTSSRKENDIVQILSGVYKGKTTGTPLSLQILNTDVDSS
PYENSERLYRPGHSQYTYEKKFGIVDPNGGGRSSARETACRVAGVVAEKFLANQNIFTL
AYLSSLGSTLPHYLKISPELIHKTPHSPYSPLPNEKIQEIATSLHDDDSDSLGVISFI
TSPIHDFLGEPLFGKVHALLASALMSIPAAKGFEIGKGFASAQMRGSQYTDFFVMEGENI
TLÄSMNCGGTLGGITIGVPIEGRIAFKFTSSIKRPCATVTKTKKETTYRTPQTGRHDPCV
AIRAUPULPAMMINIALDLIVLODESKI AIRAVPVVEAMINLVLADLVLYQRCSKL

CPh...1038 1192750 1192199

CPn_1038 1192750 1192199
areli-Shikimate Kinase II
WKLEI_RNVMTIILCGLPTSGKSSLGKALAKFLNLPFYDLDDLIVSNYSSALYSSSAEIYK
AYGDEKFSECEARILETLPPEDALISLGGGTLMYEASYRAIQTRGALVFLSVELPLIYER
LEKKGLPERLKEAMKTKPLSEILTERIDRMKEIADYIFPVDHVDHSSKSSLEQASQDLIT

CPn_1039 1194011 1192665 aroA-Phosphoshikimate Vinyltransferase VCFTMLTYKVSPSSVYGNAFIPSSKSHTLRAILWASVAEGKSIIYNYLDSPDTEAMICAC KQMGASIKKFPQILETVGNPLAIFFKYTLIDAGNSGIVLRFMTALACVFSKEITVTGSSQ LQRRPMAPLLQALRNFGASFHFSSDKSVLPFTMSGPLRSAYSDVEGSDSQFASALAVACS LARGPCSFT1IEPKERPWFDLSLWWLEKLHLPYSCSDTTYSFPGSSHPQGFSYALDVACS
LAEGPCSFT1IEPKERPWFDLSLWWLEKLHLPYSCSDTTYSFPGSSHPQGFSYAVTODFS
SAAFIAAAALLSKSLQPIRLRNLDILDIQDKIFFSLMQNLGASIQYDNEEILVFPSSFS
GGSIDMDGC1DALPILTVLCCFADSPSHLYNARSAKDKESDRILAITEELQKMGACIQPT
HDGLLVNPSPLYGAVLDSHDDHRINALTIAALYASGDSRIHNTACVRKTFPNFVQTLNI
MEARIEECHDNYSMWSTHKRKVFARESFG

1194876 1194073

NO robust homolog present in Genebank/EMBL as of 11/7/98
RPSGSLFLRTWSPSSSFREHTVCAAPLLYPRRRSPDYLFSPTGCPMSTTTVKHFIHTAS
RPSGSLFLRTWSPSSSFREHTVCAAPLLYPRRRSPDYLFSPTGCPMSTTTVKHFIHTAS
REPVLKEIVASNYWHAQWINTLSFLENSGAKKISASEHPTEVKEEVLKHAABEFRHGHYL
KTQISR ISETSEPDYTSKNELGGLITKYYIHLDLBRTGRVLENEYSLSGGTLKTAAYILV
TYALELRASELYPLYHDILKEAGSKITVKSITLEEQGHLQEMERELKDLPHGEELLGYAC QFEGELCLQFVERLEQMIFDPSSTFTKF

CID_1041 119630 1 1191726

*bioA-Adenosylmethiotice 8 Ameno-7-Oxonorangle

Ameno-tameters 8

DERRIE FULLIERFUSIK VETEVMFLORUSES LORK ZOHLEFKORANALTY POYLRYGGK

WERRIERALANDERO, YMPODOQUKRNHKDNIKLLOPTMEDSMONGGO IMPET

OSALDOTPER IVEGENAYLYA ELSTER ZLOALSSWONLIGIKHEY TIKKLOEOAOKLEHV

LPANETHEEN ELM SKLAFILEFESGLER FESDAC, TOLIE IAMY TAVQYYYNGHKAKSHEV

GENNAYLEFTEM EASTE VETEVEREBLE LEGST LAAPYZOKE FLA LAQAKTYPSESNI

AAFTYEPILA MAXIMIMYNPEGLKE LEKLAKHYOTZ TAOETTAPYZOKE FLA LAQAKTYPSESNI

FEDET CLESK EZEKYLEFA LEPTTRE HID NYVQODMKALLIFERTET NPLOGSAALASL

DEFLOERET ACROMITER ERGEFORANGEMORG CVID TYLALDYDALATGYFSQYRDHLN

REFLERGVLLRPLGNTLYVLPPYC IQEEDLR I IYSHLQDALCLQPQ

1196629 [1195934

CPn_1042 119639 1195334 *bloD-dethloblotin synthetase NRSPFTYFRANFFMORILIVGIDTGVGKTIVSAILARALNAEYWKPIQAGNLENSDSNIV HELSGAYCHPEAYRLHKPLSPHKAAQIDNVSIEESHICAPKTISNLIIETSGGFLSPCTS KRLQGDVPSSWSCSWILUSQAYLQSINHTCLIVVEAMRSRNLNILGMVVNGYPEDEEHWLT QEIKLPIIGTLAKEKEITKTIISCYAEQWKEVWTSNHOGIQGVSGTPSLNLH

75 3 11 14 573

DIOF_2-OXONONANOALE SYNTHASE_2
PMLCOOFLIEALARRKSKHTYRSLSLNSHLIDFTSNDYLGFASSPELRKEYITKLHAIES
LGATGSRLLTGHSQLCQRIEEQLAAYHNFESCLIFNTGYTANLGLLYALATDQDRILHDL
YIHASIYDGIRLSKAQSFFPNHNDLNHLEKRLASSHLGRTTVCVESVYSLHGSVAPLQAI
SELCERYSAYLIVDEAHAVGVFGDQGGLVSALGLQDKVLATVYTFGKALGTHGAAIAGS
SILKDYLINFCRPFIYTTAQPPHALTAIELAYEHNQRAFNQREHLSALIHHFREKAQNLG
LQLMKDNTTTPIQSICVSGSHRARQAALQIQNSGYDVRPIVSPTVKQREELLRICLHAFN TKNEIDHLLHTLEQIFLCNVSSL

1198700 1197699

*bioB-Biotin Synthase

*blob-Blotin Synthase
AKHAREETVSWSLEDIREIYHTPVFELIHKANAILRSNFLHSELQTCYLISIKTGGCVED
CAYCAQSSRYHTHVTPEPMMKIVDVVERAKRAVELGATRVCLGAAWRNAKDDRYFDRVLA
MVKSITDLGAEVCCALGMLSEEQAKKLVDAGLYAYNHNLDSSPEYETIITTRSYEDRLN
TLDVVNKSGISTCCGGIVCMSESEDDRIKLHVLARTRDHIPESVFVNLLWFIDSTPLQDQ
PPISFWEVLRTIATARVVFPRSMYRLAAGRAFLTVEQOTLCFLAGANSIFYGDKLLTVEN NDIDEDAEMIKLLGLIPRPSFGIERGNPCYANNS

1199602 1198901

CPH_1045 "conserved hypothetical bacterial membrane protein GTLPMNTSHRKTLVFSYLSSTFTLLLVLSNLVLSSKLIPTTFFNFIIPGGLILYPLTFLI SDVVNEIFGPKKARVMIFSAFIANLLASSIVQIFMFFPVASPEMQTAMHCLFDLSPLRFL ASLLAFIVSQOLDIVLYTFFKNRTPNSSLWIRSNGSTWISQIPDTFIVDTCILYFGMGLS FPQTLNIMFYSYIYKITFCVLTTPLFYLAVNTIRKFLGMPSTKIANTVPLINQP

1200675 1199590

**TTYPTOPHAN HYTOXYIASE
VHYCERTLDPKYILKIALKLRQSLSLFFONSQSLQRAYSTPYSYYRIILQKENKEKQALA
RHKCISILEFFRNLFVHLLSLSKNQRBGCSTDMAVVSTPFFNRNLWYRLLSSRFSLMKS
YCPRFFLDYLEAFGLLSDFLDHQAVIKFFELETHFSYYPVSGFVAPHQYLSLLQDRYFPI
ASVMRTLDKDNFSLTPDLIHDLLGHVPWLLHPSFSEFFINMGRLFTKVIEKVQALPSKKQ RIOTLQSNLIAIVRCFWFTVESGLIENHEGRKAYGAVLISSPQELGHAFIDNVRVLPLEL DQIIRLPFNTSTPQETLFSIRHFDELVELTSKLEWMLDQGLLESIPLYNQEKYLSGFEVL

CPn_1047 1200537 1201343 dapB-Dihydrodipicolinate Reductase FGSRNMGSSSHVOVIOCSGRTGKVIVSALEQSSEYTLGPGFSRSSALTLFQVIAHNDVLV DFSHPLLTKEVVAHLLISPKPLIIGTTGFPGKCKEAHDSLEELTHIVPVVVCPNASLGAY IHRRLVMLLSQLCNPQFDIRIRETHHRYKKDSLSGTAQDLLDTIQQVKQEDWGEEYEVGQ RDSSKKTIEVQSSRVGDIPGEHEVAFISSGEQILVRHTVFSRNVFGRGILSILDWLKTLN PQPGLYSLGDTLELVLRNEHCLLKKTTDH

1201588 1202604 CPn 1048

asd-Aspartate Dehydrogenase
DGERKGMRIAVLGVTGLVGQKFVALLHKWYRDWVIAEVVASNSKYGQSYGDACIWQEPIG DGERGGRITAVAS VGLVQAFVALLIRMYKLWY ILEVVASNIKYQGYGGL WQEFIG
PMPEMVRDLPIRKIEEVQSDIVVSFLPSSAESMEAYCLSQGKVVFSNASTYRMHSSVPII
IPEVNSDHFQLLEEQPYPGKIITSPNCCVSGITLALAPLRKFSLDHVHIVTLQSASGAGY
PGVPSLDLLANTVPHIVGEEKILRETVKILGSSKOPLPCKLSVTVHRVPVAYGHTLSLH
VTFSKDVDLDEILYSYQEKNKEFPNTYQLYDNPWSPQARKHLSHDDMRVHLGPITYGGDF
RTIKMNVLIHNLVRGAAGTLLASMENYFFDYLKREMCLR

1202586 1203914

LYSC-ASpartokinase III
EGNYSKIYYKFGGTSLATAENICLVCDIICKDKPSFVVVSAIAGVTDLLVDPCSSSLRER
EEVLRKIEGKHEEIVKNLAIPFPVSTWTSRLLPYLOHLEISDLDFARILSLGEDISASLY
RAVCSTRGWDLGFLEARSVILTDDSYRRASPNLDLMKAHWHQLELNGPSYIIQOFIGSNG
LGETVLLGRGGSDYSATLIAELARATEVRIYTDVNGIYTMDPKVISDAQRIPELSFEEMQ NLASFGAKVLYPPMLFPCMRAGIPIFVTSTFDPEKGGTWYYAVDKSVSYEPRIKALSLSQ YQSFCSVDYTVLGCGGLEEILGILESHGIDPELMIAQNNVVGFVMDDDIISQEAQEHLVD VLSLSSVTRLHHSVALITMIGDNLSSPKVVSTITEKLRGFQGPVFCFCQSSMALSFVVAS ELAEGIIEELHNDYVKQKAIVAT

CPn_1050 1203884 1204798 dapA-Dihydrodipicolinate Synthase LCKTKSYSRHVGRIMHLLTATVTPFFPNGTIDFASLERLLSFQDAVGNGVVLLGSTGEGL SLTKKEKQALICFACDLQLKVPLFVGTSGTLLEEVLDWIHFCNDLPISGFLMTTPIYTKP KLCGQILWFEAVLNAAKHPAILYNIPSRAATPLYLDTVKALAHHPQFLGIKDSGGSVEEF QSYKSIAPHIQLYCGDDVFWSEMAACGAHGLISVLSNAWPEEAREYVLNPQEQDYRSLWM ETCRWVYTTTNPIGIKAILAYKKAITHAQLRLPLSIEDFDLENVSPAVESMLAWPKLRTS

CPn_1051 1201956 1205270
No robust homolog present in Genebank/EMBL as ot 11/7/98
FFMTPMSIQQLHLIKTIDPVRKISPVTTYKSSFROSLLRFLELFWMFLYCIRSIRFHCV
HIATFICRGLILFLTTLFLSMICILHFITLPWICKEDPRIIRKNK

CPn_1052 1205402 1206169
No robust homolog present in Genebank/EMBL as of 11/7/98
FFICKMKYNSREKIKSALRICGOYGITYFRNNIGLGGYDKIFYGLGGYYFROPNSIGRGR
GECFFRIKKTEVETREKVKIKDE IPPGLGONDFYKVAESFPKRHAALEGLGSGOSIGNLCA
LONFLDGAMLGRNFCKEIMGSTIFTRIKOTTPAGGSEPFRYTAGGYLAGHROKLAGGYEL
GYTAGLLGGRLKDVJDGHRTRITTGTLGVHGGMYTRPLGCTKYIVGKARPLLFFFRLTSD VRPOLKKKERLLUCKO

(Ph_10%) 120c138 (20670) No robu it homotog pre em im demolank/EMBL e, of 1177798 KK SDLAF-AKTOINBLYDT (LABL/WACFILL-TIX "LENYSSEMAJBEVLVY" KERC (19GEP BLAT WINDTYY) UVILE IGLÆ ZEYTSE EGBRIÐEN I DMIMARK (LAVLSBETREAKETER SPDYAFFSUTARESUMI SEKLAMTFOVSEV LONGY SOCTYVTKTNEKEOYRHE SHMTGF

CPn_1054 1207010 1207466
No robust homolog present in Genebank/EMBL as of 11/7/98
SRWHRFIMQVLLSPQLPPPPOHSVGSISSPSKLRVLAITFLVFGMLLLISGALFLTLGI
FRLSAAISFGLGIGLSALGGVLMISGLLCLLVKREIPTVRPEEIPEGVSLAPSEPALQA
AQKTLAQLPKELDQLDTDIQEVFACLRKLKDSKYESRSFLNDAKKELRVFDFVVEDTLSE IFELRQ IVAQEGWDLNFLINGGRSLMMTAESESLDLFHVSKRLGYLPSGDVRGEGLKKSA KEIYAPLMSLHCEIHKVAVAFDRNGYAMAEKAFAKALGALEESVYRSLTQSYRDKFLESE KETYAPIANSIAGETHKVAVAFORNSYAMAEKAFAKALGALEESVYRSLTGSYRDKFLESE

KERMI BETTÄNE JUJA – KATALEMALHAMKE HER AVTAVITEUR JED JOHNET

EL HER LEMBETTUH DA FETTENBULHEN ALLAMTENER PETKINGA ZETANAPPLY

TVRDMYDQEFQKAGERLEKLHALYPEVSVSTRENKTQETRSNLEKAYEATEENYRCCVRE

OEDVYKEEEKREAFRERGNKTLSPEELESSLEOFDHGLKNFSEKLMELBGHTLKLQKEA

TAEVENKTLSDAESRLETVFEDVKEMPCRTEETEKTLRMAELPLLPTKKAFEKACSQYNS

CAEMLEKVXPYCKESLAYVTSKERLVSLDEDLRRAYTECQKRFGODSGLESEVRACREOL

RERTIGEFETGGLDLVEKELLGVSSKLRNTECDCVSGVKKEAPPGKKFYAQYYDETYRVRV

QSRMTMSERLREGVQACNKMLKAGLSEEDKVLKEEFYWLYREERKNKEKRLVGTKIVAT

QQRVAAFESTEVPETPEAPEEKPSLLDKARSLFTREDHT

1209583 1210521

CPn_1055 1209583 1210521

No robust homolog present in Genebank/EMBL as of 11/7/98

CKYLYHHSYPPPPHSVGAFFCLSKFRVLAITFLVLGVLFLISGALFLTLGISGLSAAIS
FGLGIGLSALGGVLVVSGLLCLLAKREVPTVRPEEIPEGVSVAPSEEPALQATQKTLAQL
PKELDQLDRYIGEVVSCLGKLKDLRCEDGLLKDAKEKLQVFDFVWKDMMTEFVELQQIM
DQEGMYLKCLIQEMRDIGSTLFMSQVSLFKLWEBKLGVLPSGDVRGERLKKSAREVVDRFM
RRICDTRKVAMTFDRNAYGVAKTAFEKAFGALETCVYKSMTESYREAFCEYKKTKILRDE
EKILRICYLELRR

CPn_1056 1210482 1211228
No robust homolog present in Genebank/EMBL as of 11/7/98
GEDIKDMLSRVEEIEMMLRVIELPLLPIKQALEKAFVQYNSYKAKLTKVEPCFRESPAYI
TSEERLQSILOTILERAYKEYQKRFQEPSRLESEVSGCREHLREQVKQFETQGLDLIKEEL
IFVSDVLFRKMVSCLVSTVHVPFMEFYYEYFELHRLRAQMMANAEIYSKVRKAFPEML KETLEKAKAPREEEYWLLCEERKSKEKRLILNKIEAAQQRVKDLEPPPIKETGKOKRKKE

CPn_1057 1211467 1213596

CT356 hypothetical protein
IIHFYFFNFAMPEPLYTNKLITEKSPYLLLYAHTPVNMYPMGAEAFHIAAIENKPVFLSI
GCKHSRWCQVMLQESYTNPBIAAMINEYFVNVKVDKEELPYVAKLYGDLAQMLAVSGDHQ
ETVSMPLNVFLTPDLVPFFSVNYLGNEGKIGLESFPQIIDKLHFMMEDAEEREALVEQAM
KVLEIASFLEGCVRKEILDESSLKRTVAALYQDIDPHYGGVKAFFKRLPGLLLQFFLRYS
LEYÖZESRGLFFVDRSLSMVALGGVRDHIGGGVYSYTIDDKWLIPAFEKRLIDNALMALNY
LEYÖZESRGLFFVDRSLSMVALGGVRDHIGGGVYSYTIDDKWLIPAFEKRLIDNALMALNY
LEYÖZESRGLFFVDRSLSMVALGGVRDHIFUPVREIELSEKYHRSIEAIEDIVDRSRDI
LEYÖZEGARSHFSKDDLSLTFNNGMNIYTFAYAGRLLGEVEYIELGKKCGEFVRNSLYKHH
ELYARWREGEAKYRASLEDYGALILGVLALYESGCGSFWLSFAEELMQEVVLSFRSEEGG
FYSVJGRDSTLLIKQSPLSDGETISGNALICQCLLSHLITEKKHYLTYAEDILQIAQAC
AHTHKFSSLGLLIASGNYFSRKHVKVLIALGDDENSSPUKCLSGLFPYLSLIMMTQEN
QEHLETVLPEYEHCLIPKGDCTATTIYVLEVDQCKRFKDLELFRRYLISL

CFPLIO58 1213742 1214836

CPN_1058 1213742 1214836
CTPS5 hypothetical protein
EVMKLYOTLRGIVLVSTGCIFLGMHGGYAAEVPVTSSGYENLLESKEQDPSGLAIHDRIL
FKVDEENVVTALDVIHKINLLFYNSYPHLIDSFPARSGYYTAMWPVVLESVIDEFLMYAD
AKAKKIATDPTAVNQEIEEMFGRDLSPLYAHFEMSPNDIFNVIDRTLTAQRVMGMMVRSK
VMLKVTPCKIREYYRKLEEEASRKVIWKYRVLTIKANTESLASQIADKVRARLMEAKTWD
KDRLTALVISQGGOLVCSEEFSRENSELSOSHKQELDLIGYPKELCGLPKAHKSGYKLYM
LLDKTSGSIEPLDVMESKIKQHLFALEAESVEKQYKDRLRKRYGYDASMIAKLLSEEAPP

CP#=1059

CPE_1059 1214848 1215678
kgsA-Dimethyladenosine Transferase
vTRSSPAQUSRFLSEIQNKPKKSLSQNFLVDQNIVKKIVATSEVIPQDMVLEIGPGFGAL
TEELIAAGAQVIAIEKDPMFAPSLEELPIRLEIIDACKYPLDQLQEYKTLGKGRVVANLP
VHITTPLLTKLFLEAPDFWKTVTVAVQDEVARRIVAQPGGRDYGSLTIFLQFFADIHYAF
KVSÆSCFYPKPQVQSAVIHMKVKETLPLSDEEIPVFFTLTRTAFQQRRKVLANTLKGLYP
KEOFTQALKELGLLLNVRPEVLSLNDYLALFHKMQAG

CPn_1060 1217694 1215727

CPm_1060 1217694 1215727

dxs/tkt-Transketolase
YKRPLYIHITKVMTSSSCPLLDLILSPADLKKLSISQLPGLAEEIRYRIISVLSQTGGHL
SSNLGIVELTIALHYVFSSPKDKFIFDVGHQTYPHKLLTGRNNEGFDHIRNDNGLSGFTN
PTESDHDLFFSGHAGTALSLALGMAQTTPLESSTHVIPILGDAAFSCGLTLEALNISTD
LSKFVVILINDNNMSISKNVGAMSRIFSRWLHHPATNKLTKQVEKWLAKIPRYGDSLAKHS
RRLSQCVKNLFCPTPLFEQFGLAVVGPIDGHNVKKLIPILQSVRNLPFPILVHVCTTKGK
GLDQAQNNPAKYHGVRANFNKRESAKHLPAIKPKPSFPDIFGGTLCELGEVSSRLHVVTP SLOQAQNNPAKYHGVRAND NAKESAKHLDAIKEKPSFPDIFGGTLCELJEVSSRLHVVTF AMSIGSRLEGEFKQKFPERFFDVGIAEGHAVTFSAGIAKAGNPVICSIYSFTLHRALDNVF HDVCMQDLPVIFAIDRAGLAYGDGRSHHGIYDMSFLRAMPQMIICQPRSQVVFQQLLYSS LHWSSPSAIRYPNIPAPHDPLTGDPNFLRSPGNAETLSQGEDVLIIALGTLCFTALSIK HQLLAYGISATVVDPIFIKPFDNDLFSLLLMSHSKVTIEEHSIRGGLASEFNNFVAFFN FKVDILNFAIPDTFLSHGSKEALTKSIGLDESSMTNRILTHFNFRSKKQTVGDVRV

CPn_1061 1217932 1217666 CT330 hypothetical protein FGSLMVEIHHKDPSLKKLFALOGSLETLNSLSDIVATYEAMFSLIYEGLNKALRKDQLCY LLSVNSKGELLKSPSGDPIVQTFPIHPHH

1219835

CPh_1062 1219835 1218150

xseA-Exodoxyribonuclease VII

RGFPMD3PPQAVASUTER INTLLESNFCQIIVKGELSNVSLQPSGHLYFGIKDSQAFLN
GAFFHEKSKYYDEKPKDSDAVIIHGKLAVYAPRQQQIVAHALVYAGEGDLLQKFEETKR
RUTAEGYFATEKKPLPFAPQCIGVTTSPTGAVIQDILRVLSRRARNYXILVYPVTVQGN
GAAHELGKAIEVMDAEMLADVLIIARGGGSIEDLWAPNEELIVKAIHASTIPIVSAVGHE
TDYTLCLFASDVFAPTPDAAACIVCKSSEEQVQVFEGYLRHLLSHSRQLLTSKKQQLLPW
RRFLDRAEFYTTAQQQLDGIETAIQKGVOKIHESKQRYDNISRWLQGDLVSPMTCRLQS LIKKMEJQALDIKALDOVRCHQLKKOLTYPROTOQAJDKEJPWRQQLDTLEJPREHYQKE EYFHKHTPLKHAHNVLEQQLRSHVQKLELLGPREJPCCELNEQNQKTALANVKCTLATIL EHRYENDVARYDALKEQEHJENPKNVLKRGYAMEDDTHENDAMEDVDSLQEHARVRTQLQ DEVELLALATION OF TORPING

1219900 1220712 CPn 1063

EPIS-Triosephosphale Isomerase
FCRESMRIKFRENKERKMTRQSYVLGNWKMHKTIQEAKEYVQTLAJLLQGEFLJGTIGIA
SPETSLRAIHEVINTTGAFLWLGAQNVHPELSGAFTGEISLPMLKEVJVEFVLVGHSERR
HIFGESDAFIASKVKSVAQAGLUPVLCVGESLEVREGKAHQVIKKQLLGLEQMONGSE
FLIAYEPVWAIGTGKVAEASDVQDIHMFCREVVAERFSEATAEEISILYGGSVKVDNAQR
FGQCSDVDGLLVGGASLEGQSFFEVAKNFNV

CPn_1064 1220716 1220895 Note that the control of the first of the the MR. The EMPORTED SECTION OF THE CAMPACITY OF THE TOTAL THE CPn_1064

CPn_1065 1221140 1220928
No robust homolog present in Genebank/EMBL as of 11/7/98
RHRIGRHRRTSDPCFLFYFSIPEESLPPDSCRLNQMPKHEHLPSILLKKPIIDYLKITSI YEKAIFNIGLP

CPn_1066 1221132 1221488
No robust homolog present in Genebank/EMBL as of 11,7/98
SMSLNKEIGMTVLFYAFLFIFLELCVILCGLILVQESKSMGLGSSFGVDSGDSVFGVSTP
DILKKVTSWCAVAFCIGCLLLSFSTNLLGKKLDAKEFLLPAAEESDTQASSESVEADES

1221675 1222292

CFT_105/
def-Polypeptide Deformylase
IQVLVVRDFFTELCOAHVQTMIRRLEYYGSPILRKKSSPIAEITDEIRNLVSDMCDTMEA
HRGVGLAAPQVGKNVSLFVMCVDRETEDGELIFSESPRVFINPVLSDPSETPIIGKEGCL
SIPGLRGEVFRPQKITVTAMDLNGKIFTEHLEGFTARIINHETDHLNGVLYIDLMEEPKD PKKFKASLEKIKRRYNTHLSKEELVS

1223267 1222365

rnhB-Ribonuclease HII

MSCMPPPFVVTLTTSAQNNLRDQLKEKNFIFSQPQNTVFQARSNTVTCTLYPSGKLVIQG KGSEEFIEFFLEPEILHTFTHARVEQDLRPRLGVDESGKGDFFGPLCIAAVYASNAEILK KLYENKVODSKNIKDTKIASLARIIRSICVCDVIIIYPEKYNELYGKFONLNTLLAMAHA TVINNLAPKPAGDVFAISDQFAASEYTLLKALQKKETDITLIQKPRAEQDVVVAAASILA RDAFVQSIQKLEEQYQVQLPKGAGFNVKAAGREIAKQRGKELLAKISKTHFKTFDEICSG

CPn_1069 1223507 1223941 yfgA-HTH Transcriptional Regulator VIMQEHIHKELLHLGEIFRSSRESQSLSLKDVEAATSIRYSCLEAIEQGCLGKLISPVYA QFIKKYATYLGLDDDSILQEHPYVMKIFKEFSDHRMEMLLDLESMGGRNSPERAIHSWS NLWWAGLIIIGGIMVWWLGSLFSIF

CPn_1070 1225523 1224144 No robust homolog present in Genebank/EMBL as of 11/7/98 WDDDEIERDGIVTGGGFGIPCQCLRCWRKLPTEKRPNRWL

CPn_1071 1227336 1225885

No robust homolog present in Genebank/EMBL as of 11/7/98

KGTTMVCPNNSWFRMCGNFNCEWWEVTTTEETTRQSASDISEEAGSSGGAAPITTQPTKI

KVEKKRVQFNTAQGDESTIHMIQEAGELVDSILSHRRTQGCTEYCYDSYATGCQQRCGSF

GRLICGTYKACCLDREDNQVAGLVHECEQTHGPIAVALAAKTMGLNIMELVEKNTILSEE

QKNEFRQHCSEAKTQLYGTMQSLSQNFFLEGVNSIBERGLDDSLVQAVLSFIATRSWEKT

LESEEASGTSSASNSTRIPACYILINTSPLTTSFLSCGSRDARRPSSVGAEPQVVAKKYND

NGMARQLGKIQVTNLKTGDFSALGPFGLLIVKMLNSFLLSASQSTSSILKHTGGEICYTC

PNFRDIVVLLMLAIGYCPANTDETSVVDIHMIDDPIMTIFYRLQYSYRTGKTSASFLKKK

PSLVROESLDCTPPAEAVPLMSSLEEEDENENDDDECNILAYQORILEGSGHLOTLFGIK PSLVRQESLDCPTPAESVPLMSSLEEEDENEDDDEDGNLAYQQRILECSGHLQTLFLGIK

CPn_1072 1227924 1228835 No robust homolog present in Genebank/EMBL as of 11/7/98

NO IODUST INDUIGUOG PIESENT IN GENEDANK/EMBL AS OT ITUSKKEVILHANMCCWKQMLKIQKKRMCVSVVITVGAIVGFFNSADAAPKKKKIPTQILYSFT KVSSYLKNEDASTIFCVDVDRGLLQHRYLGSPGWQETRRRQLFKSLENQSYGNERLGEET LAIDIFRNKECLESEIPEQMEAILANSSALVLGISSFGITGIPATLHSLLRQNLSFQKRS IASESFLLKIDSAPSDASVFYKGVLFRGETAIVDALSQLFAQLDLSPKKIIFLGEDEVVQAVGSACIGWGMNFLGLVYYPAQESLFSYVHPYSTATELQEAQGLQVISDEVAQLTLNAL

CPn_1073 1229011 1229832
Predicted OMP (CT371)
MRRYLFMYLALCLYRAAPLEAVVIKITDAQAVLKFAREKTLVCFNIEDTVVFPKQMVCQS
AWLYNRELDLKTTLSEEQAREQAFLEWMGISFLVDVELVSANLRNVLTGLSLKRSWVLGI
SQRPVHLIKNTLRILRSFNIDFTSCPAICEDCWLSHPTKDTTFDQAMAIEKNILFVGSLK NGQPMDAALEVILSGISSPPSQIIYVDQDAERLRSIGAFCKKANIYFIGMLYTPAKQRVE SYNPKLTAIQWSQIRKNLSDEYYESLLSYVKSK

RNA JECTION

EmRNA. 133193 139074

Ribonus lear e F RNA 607342 607649

LG J EPNA 1000564 1002117 23S FRNA 1002415 1005278 50 rRNA 1005393 1005509

20

25

5

What is Claimed is:

- 1. An isolated nucleic acid encoding a *C. pneumoniae* protein as set forth in Table 3.
- 2. The isolated nucleic acid of Claim 1, wherein said nucleic acid has a nucleotide sequence of an open reading frame in SEQ ID NO:1.
- 3. A probe comprising a hybridizing fragment of an isolated nucleic acid according to Claim 2.
 - 5. An isolated nucleic acid that hybridizes under stringent conditions to the nucleic acid sequence of Claim 2.
 - 6. An expression cassette comprising a transcriptional initiation region functional in an expression host, a nucleic acid having a sequence of the isolated nucleic acid according to Claim 1 under the transcriptional regulation of said transcriptional initiation region, and a transcriptional termination region functional in said expression host.
 - 7. A cell comprising an expression cassette according to Claim 6 as part of an extrachromosomal element or integrated into the genome of a host cell as a result of introduction of said expression cassette into said host cell, and the cellular progeny of said host cell.
 - 8. A method for producing a *C. pneumoniae* protein, said method comprising:
 - growing a cell according to Claim 7, whereby said *C. pneumoniae* protein is expressed; and
- 30 isolating said *C. pneumoniae* protein free of other proteins.

- 9. A purified polypeptide composition comprising at least 50 weight % of the protein present as a *C. pneumoniae* protein comprising an amino acid sequence of claim1.
- 5 10. A monoclonal antibody binding specifically to the polypeptide of Claim 9.

10

ABSTRACT OF THE DISCLOSURE

CHLAMYDIA PNEUMONIAE GENOME SEQUENCE

C. pneumoniae genome sequence and analysis of the encoded polypeptides and RNAs are provided. The C. pneumoniae gene nucleic acid compositions find use in identifying homologous or related proteins and the DNA sequences encoding such proteins; in producing compositions that modulate the expression or function of the protein; and in studying associated physiological pathways. In addition, modulation of the gene activity in vivo is used for prophylactic and therapeutic purposes, such as identification of cell type based on expression, and the like.

SF 1040660 v1

Contig463

Length: 273254..

4

ATTGTTCCTG TAAGAACACT TCCAAAGCGC ATTTAATCAT TTTTAGTAAA AAATAAAAAT ATACTTTTAA ATGTTGAGAA AATTTTTAGC TAAACTTTAT 51 101 AAAGGGTTGT TGGTGAAACC TTTGGGTTAC TCCTCAGAAC GACTTTGTGA TTCTATAGTA TTAAAAGGAT CTTGGAGTAT AACAAGTAAA GATCTTTGAG 151 201 GATAGCGTAG GGCCGTATTT TGAATAGCGT CCAATAAAGC GCGTTTGCAA AACGCTTGAG TTTGGTTGTC CCAATAGAAA GTGCCTTCTT TAGGAAGAAT 251 CTCTTCTGGA GGCACTTCAT AGACCGAAGT AAAGAGAGGA AGAGCAACGA 301 TTGCTGCATG ACTTTCTATA GCTGCTTTAA GGCAGTTCTC GTACGCTAGT 351 AAAGCTTGGC GATAATATTC TTGCTGATTT GGTAACTCTT CAGATTTAGG 401 GCCGCATACG TGGCCTAAAA AGGTCGGAAG AATACTTTTC TTTTCTGCAG 451 AGCTTAAATT TAGATTAAAC GTTTGATCTA GAGCTTCGTT TGGAAGTTTT 501 ACTACTCTCA CTTCGGTAGG GGAAAAGGGG TCTTCTCCTT TTGCGGGACC 551 601 CCCTTCGCGT TGCTTGCATG TATCCCACAC GCTTTTATCT TTTAGGGTGG 651 AGTAAAGGAT AGTAGAGAGG TTCGTTGCAG TGTTGTCGAT CAGATTCGTT 701 GGGCCTACGG GATTAAAGAT GATCCCTGTG GATTGATTTT TTTCGATCAC 751 TCTAAGTCCA GTTAAGAAAG TAGGCTGAAA TGGTTGAGAC GCATCTGTTT 801 GTATCGCTAC CTTGAACTTA GGGTTCAGGT GATTATTGTA AAATTGCATC 851 TCGTTTGAGT AGCAGTCTAC GTTTTTTTCT TGCCACGCTT TTCCCAAAGG CTTGAAGTTT TGCTCTAGAA CTTTCTGCCA GTTAGAAGAT ACCTTTGAGG 901 951 TCATTTGGTG GTAGACTAAG AAGGTTACAA CTGAGAAGAG GGCCGTGGTA ATGAGAAGAG CCAAAAATAC AGGGTTCCCT AATACTATCG TTAAAGAGAT 1001 TCCAGCCACC AAAGCTCCTA AAGCTAAAGA AGCTAGGATT GCAAGAGTGG 1051 ATATTTTGC TATGGTAAAC TGTTTTTTAG GAGCAATTTC TTTATCCCGA 1101 1151 GGCACATAGG ATAGTACAGA AACTTGAGAG CTCTCAGTAC GTGAGGGTCC 1201 TGACATAACA TTTTTTTTGT AAAATACTTT CTATAATTTT AACATATTTG 1251 TGTTTATCGA TCCGAGAAAA TTGGAGAGTG AGAGCGCATG TCTTGCAATT

1301 TAGAATGATC GGGGACGACA TCTAGAGCTA TGTAGACATT GCGTGCGTAG 1351 TGGGAGCAAA TATAGCGAGA TATAAAGTAT AAGGGAATTG CTGTTAGGAA 1401 GATAAAGGAG CACAAAGGGT GGATACATAG CCCAATAGCT ATGGTGGTAG 1451 CAATCAGAGC TATCCAGACG AGTGCAATCG CAATAGTAAC GAAGAGGGCA 1501 AGCTTGAAAT TATAGCGAGG ACGAGTAGCT GGGGGAAATA GAGAGGGAGC 1551 CGTTCCATCA AAACCGGGAG TAGCTGAAGA AGCCATAAAC TATTAAAAAT 1601 TAAGTTTTTT TCGGAGCATA AAGCATTTTA AAGTAGTGGG GTCTTTTTTG 1651 TCACGGAGAT GTCCTGGACT TCCCAAGCGT TTCTAACAAA GATACCTGCT 1701 TTTGAGAGGA GAACTTTTGA AACTCCTGCA AGGTCATCCT TCCTTGGCAC 1751 CAGTAGGTTT TTTCAGGAAA TCGCGGAAAG ATTTTGGCGA AAGCTCTTAC 1801 AGTTGAAGGG CTTGTGAAGA TAATTTTTTT GTATTTAGAT AAAATATTTT 1851 TTTTAAGTTT TCGCGGCTTC ACTGTGTAGT GAGGGTAAGA GAAAAAAGTA 1901 AATCGATTGT AAAGAAATTC TCTGATCACA GGTCTTGCGA GGGAGGAGTG 1951 GGGGTAGAGA ATGCGGGCTG AAGAGGGCAG TGCCTGTAGC AATGGGAAGA 2001 TGCCTTCAGC GATTTCTTGA GTTGCTACTA CGTACTTCAC TTGTCCAAGG 2051 AAAGAGAGA GTCTTTCTTT GGTGGACTCT CCTATACAGA GGTAGGTCTT 2101 TGTTTTTAGA GTGGCCTTAG AAAGAAGAA AGTCATTCTG GAAAGGAATA 2151 GGTGAGTGGA TGAGGGACTT GTGAGAATCA CATGGGTTGC TTGTGGAAGG 2201 AATTGAAGAG CACGCTTATT TTGTGGAGTG CTTTTTGCAT AGGGGAAGAG 2251 AGTTAGAATA GGCAAATAAT GAGCTTGGTA TTTACGAGCG GTTTTTTGAT 2301 TCAATCCTAA GTAGAGGGTC ATGAGTCTTT TCGGGTAAAA GGAAGGCTGC 2351 CTAAGTTTTT GTACCTTCAA AGGGATATAT TGAAAATAAT TTTTCTTTTT 2401 CCCTTGGTTC TTCTTGATCA TGCGTTGATT GACATTTTTC ACTTTGAAGG 2451 CTAGGCTGGT TTTTTCTGGA CTTAGAGGTT CTCTCTATTA AGGCTTCGTC 2501 TTTAGAAGTC CTTGCTAAAA GTTTTTGAGA AATTTAAGAA ATTCGCAATA 2551 GTGGAAATAT TTACAAAGGT GGTTGCGGTG GTTTCGTTGT TGCATAAGTT 2601 TTTAGAAAAT GCTTCGGGGA AAAAGGGACA AAGTTTAGCT TCGACAGCGT

2651 ATTTAGCAGC TCTTGACCAT CTCTTAAATG CGTTTCCTTC CATTGGGGAG 2701 AGAATCATTG ATGAGTTGAA GAGCCAGCGT TCCCATTTAA AGATGATTGC 2751 TTCTGAAAAC TATTCTTCAC TTTCAGTGCA GTTGGCTATG GGGAACTTGC 2801 TCACAGATAA GTATTGTGAA GGAAGTCCCT TTAAGCGTTT CTATTCCTGT 2851 TGTGAAAATG TAGATGCTAT TGAGTGGGAG TGTGTAGAGA CAGCGAAAGA 2901 ACTTTTGCT GCGGATTGCG CTTGTGTTCA GCCTCATTCT GGGGCTGATG 2951 CTAATTTACT GGCAGTAATG GCCATTCTCA CGCACAAAGT CCAAGGCCCA 3001 GCTGTCAGTA AGTTAGGTTA TAAAACTGTA AACGAATTAA CAGAAGAAGA 3051 ATACACTCTA CTTAAGGCTG AAATGTCTTC TTGTGTTTGC TTAGGACCTT 3101 CATTAAATTC TGGAGGCCAT TTGACCCATG GGAACGTACG TTTAAATGTG 3151 ATGTCTAAGC TTATGCGTTG CTTCCCCTAT GATGTCAATC CGGATACGGA 3201 GTGTTTTGAT TATGCAGAGA TCTCCCGGTT AGCTAAGGAG TATAAACCTA 3251 AGGTACTGAT CGCAGGATAT TCTTCCTATT CTCGAAGATT AAACTTTGCA 3301 GTTTTAAAAC AGATTGCAGA GGATTGTGGA TCTGTCTTGT GGGTAGATAT 3351 GGCGCATTTT GCAGGCCTAG TTGCTGGGGG AGTGTTTGTT GATGAAGAAA 3401 ATCCTATTCC TTATGCAGAT ATAGTGACAA CAACAACGCA TAAGACATTA 3451 CGCGGTCCTC GCGGGGGATT AGTTTTGGCA ACTCGAGAGT ATGAAAGCAC 3501 TCTCAATAAG GCGTGTCCTT TGATGATGGG AGGTCCTCTA CCTCACGTGA 3551 TAGCTGCTAA AACAGTGGCT TTGAAGGAAG CTCTCTCTGT GGATTTCAAG 3601 AAATACGCTC ATCAGGTTGT AAATAATGCT CGTCGATTAG CAGAGAGATT 3651 TTTAAGTCAT GGGCTACGTC TTTTGACGGG AGGAACAGAC AACCACATGA 3701 TGGTGATTGA TTTAGGTTCT TTGGGCATTT CTGGAAAAAT TGCTGAAGAT 3751 ATCTTGAGTT CCGTAGGAAT TGCTGTGAAT CGGAATTCAT TACCTTCAGA 3801 TGCTATTGGT AAGTGGGACA CTTCAGGTAT ACGTTTAGGA ACCCCTGCAC 3851 TAACGACTTT GGGTATGGGT ATCGATGAAA TGGAAGAAGT TGCAGATATT 3901 ATTGTGAAAG TATTGCGAAA TATTCGTTTA AGTTGCCATG TTGAAGGGAG 3951 TTCTAAGAAA AATAAAGGGG AACTTCCTGA AGCCATAGCG CAGGAAGCTA

4001 GAGATCGTGT TCGCAACTTG TTGCTGCGTT TCCCGCTCTA CCCTGAAATT 4051 GATTTAGAAG CTTTAGTTTA GTTAGGAGAG ACATTATTTT ATGGCAGACG 4101 GGGAAGTTCA TAAATTACGT GATATTATAG AAAAAGAGTT ATTGGAAGCG 4151 CGCAGAGTAT TTTTCTCAGA GCCTGTAACA GAGAAAAGTG CTTCCGATGC 4201 AATTAAAAAG CTTTGGTATT TGGAATTAAA AGATCCTGGA AAGCCTATAG 4251 TTTTTGTGAT CAATAGTCCT GGGGGATCTG TGGACGCAGG TTTTGCTGTT 4301 TGGGATCAAA TTAAAATGTT AACCTCACCC GTCACTACTG TTGTGACAGG 4351 GTTGGCAGCT TCTATGGGCT CGGTATTGAG TTTATGTGCA GCTCCTGGAA 4401 GGAGATTTGC AACTCCTCAT TCTAGAATTA TGATTCATCA ACCTTCAATA 4451 GGTGGACCGA TTACCGGTCA GGCAACCGAT TTAGACATTC ATGCGAGAGA 4501 GATTTTAAAA ACAAAAGCTC GCATTATAGA TGTCTATGTA GAGGCGACAA 4551 ATCAACCTCG AGATATCATA GAAAAGGCTA TCGATAGAGA TATGTGGATG 4601 ACAGCCAACG AAGCTAAGGA TTTTGGTTTA TTGGATGGCA TTTTATTCTC 4651 CTTCAACGAT CTCTAAATAT TTTATCTATT CTGGAGCAGG AAATCGTTTC 4701 CTTCTTGGTG AAACACTTCC TGAGGTTGAA GATGTTCGGT TCTTATGCCA 4751 AGAGACGAGG GTTGATGGTT TTTTATATTT AAAGCCCTCT TCTTGTGCTG 4801 ATGCGCAACT CATTATTTTT AATTCCGATG GATCACGTCC AACGATGTGT 4851 GGTAACGGCT TGCGTTGTGC GATTGCTCAC TTAGCTTCTC AGAAGGGAAA 4901 ATCGGACATC TCTGTATCTA CGGATAGTGG TCTATATTCA GGATATTTTT 4951 ATTCTTGGGA TCGTGTGCTT GTAGATATGA CTCTCGCAGA TTGGAGAGCT 5001 TCTGTTCATC GATTGGAGTC GCGTCCTGAT CCTCTTCCCA AAGAGGTCGT 5051 TTGTATCCAT ACGGGAGTGC CTCATGCTGT CGTAATTCTT CCTGAGATTT 5101 CTACTTTAGA TCTTTCTATC TTAGGTCCTT TTCTTCGCTA TCATCAGACC 5151 TTCTCTCCAG ATGGGGTGAA TGTCAATTTT GTTCAGATAC TGGGACATTG 5201 CCAGTTGCGC GTTCGTACTT ACGAACGTGG AGTCGAAGGG GAAACTGCAG 5251 CTTGTGGAAC AGGGGCTCTA GCTTCTGCTC TTGTTGTGTC AAACTCCTAT 5301 GGATGGAAGG AGTCGATCCA AATCCATACT TGGGGTGGAG AGCTTATGAC

5351	TGTGAGTCAA	AATAGGGGAC	GGGTATATCT	TCAGGGCTCT	GTAACTAGAG
5401	ATTTATAATT	AGATGTGATT	TTTGATTTTG	TCATGCAAGG	ATTTTAAAAT
5451	CTTGTTTAGG	GATAGATCTT	GCTCTCTAAC	TGGGATTTTT	CTATAATCGT
5501	AATTTATGAT	GACGTATCCT	GTACCACAAA	ACCCACTTCT	TTTAAGAATC
5551	CTTCGTCTTA	TGGATGCATT	CTCTAAGTCT	GACGATGAGA	GGGACTTTTA
5601	TTTAGATCGT	GTTGAAGGGT	ТТАТТСТСТА	CATAGATTTA	GATAAAGACC
5651	AAGAGGATCT	AAATAAGATT	TACCAAGAAT	TAGAAGAGAA	TGCCGAGCGG
5701	TATTGTTTGA	TTCCGAAGTT	GACGTTTTAT	GAAGTAAAA	AAATCATGGA
5751	AACGTTTATC	AATGAAAAGA	TTTATGATAT	CGATACCAAA	GAAAAGTTCC
5801	TTGAGATTTT	GCAATCCAAG	AATGCCCGTG	AGCAGTTTTT	AGAGTTTATT
5851	TATGATCACG	AGGCAGAGTT	AGAAAAGTGG	CAGCAATTTT	ATGTAGAGCG
5901	TTCTCGAATT	CGAATTATAG	AATGGCTTCG	CAATAATAAG	TTCCATTTTG
5951	TCTTTGAAGA	AGATCTAGAT	TTCACAAAGA	ÂTGTTTTGGA	ACAGTTGAAA
5001	ATACATTTGT	TTGATGCCAA	GGTGGGGAAA	GAAATCACTC	AAGCGCGTCA
5051	GTTGTTGTCG	AACAAAGCTA	AGATTTACTA	TTCCAATGAA	GCATTAAACC
5101	CTCGTCCGAA	ACGAGGCCGT	CCTCCGAAGC	AATCTGCTAA	GGTAGAAACA
5151	GAAACAACAA	TTTCGAGTGA	TATTTATACA	AAAGTCCCTC	AGGCTGCTCG
5201	TCGTTTCCTT	TTCTTACCCG	AGATTACTTC	ACCCTCTTCA	ATTACTTTCT
5251	CAGAAAAATT	TGATACGGAA	GAAGAATTTC	TTGCTAACTT	GCGCGGTTCG
301	ACTCGTGTTG	AAGACCAGCT	GAATCTTACC	AATCTTTCAG	AGAGGTTTGC
351	TTCTCTTAAA	GAGCTTTCGG	CTAAGCTTGG	TTACGACTCT	CTTTCTACTG
401	GAGATTTCTT	TGGTGATGAT	GATGAGAAAG	TGGTCACTAA	GACGAAGGGG
451	AGCAAGCGAG	GCCGCAAAAA	ATCTTCTTAA	TCTTCTATTT	TGTGAAGTAG
501	ATTTTTTTTA	GACGCTGTTC	TTATTGCTTC	TTTACATGAT	CTTATTACAA
5551	ATCTTTCTTA	TTTCTATTTA	TTGTTTTGTT	AAAATTTTAA	CAATAGCTAT
601	TTATTATTAG	TCATTTTTTT	AAAAAAAAA	CTGTTAAAAT	TTTTAAAGCT
651	AATTTAAGAA	ACAGTGAATA	CTTCATCATC	тсатсастас	ФС⊅ССТСССС

6701 AAGAATAGAG CCGACTCGGG TTACCTGTAG CTTAAAGACG TATCTTGAGG 6751 ATACGAGTCA GAATCAGTTG AGCACACGTC TAGTTCGGGC AAGTGTCATC 6801 TTTTTATGCG CATTGTTGAT CATTTTGGTT TGTGTGGCCC TCTCTAGTTT 6851 GATTCCAAGC ATTATGGCCT TGGCGACCTC TTTTACGGTA ATGGGGTTAA 6901 TTCTTTTGT GATGTCACTT CTTGGTGACG TTGCAATTAT AAGTTATCTT 6951 ACTTATAGCA CTGTTACGAG TTACCGGCAA AATAAGAGAG CTTTTGAGAT 7001 TCACAAGCCC GCTCGCTCCG TTTACTACGA GGGGGTCCGC CATTGGGATT 7051 TAGGACGATC ATCTTTAGGC ACAGGCGAGA TTCCTATAGT AAGGACGTTA 7101 TTCTCTCCAT TTCAGAACCA TGGTCTTAAC CATGCCTTAG CTGCTAAAAT 7151 TTTCCTATTT ATGGAGCATT TCAGCCCTGA GCCACCGAAC GAGCCTTTGG 7201 TGGATTGGGC CTGTTTGATT CGGGATTTTA GGCCTCACGT CAGTTCTTTG 7251 TGCTTTGTTA TTGAAAAACA AGGGTCATCG CTGAGGACTA AGGAAGGCAA 7301 TACGATTTGT GAGGCTTTCC GCTCTGATTA CGACGCCCAT TTTGCTATGG 7351 TAGATTGCTA CCGGTTGATC CACTCTAAGT TGATTATAGA GAAAATGGGA 7401 TTGAAGAATA TCGATATCAT TCCGAGTGTC ATGGTTCGTG AAGATTATCC 7451 TAGCCGTCCT GGGGAGGGCT ATCGCGAAGG CCTATTACGT ATGTATGGTG 7501 GCAAGGGGC TCTGTGACTT CCCTACTTTA GTTCCTAATG AGCGCTTGCC 7551 CATAGGGCCT TTCTTTGTCC CGCAGCACAC TTCCGGTGCG AAGGGTAAGG 7601 AGTTTGCTAA AAGGAATTTT TCTATAATTT CGGGATTGGA TGACATATTA 7651 AAATTATGTA TTCTTCAAAG GCGTCCTTTT GCTTTGCAGT GGGATAACCT 7701 CTCTGTGAAA AGTGATTATG AGGAGGCTGG GCCCGCTATT GGGATACGTT 7751 CTCTTGAGCC ACAAGTTTCT CAAATTTCTC CAGCCCACGG CCGGCTATGT 7801 AGTACTTTGG TCCAGTGGGC CCCTATCCTT GGTTCTGAGG AGCAGCTAGT 7851 TTGGTTAGAA GAAACAATGA AGCGCCTAAA GTTTCCTAAA AGTTTAGGTA 7901 GTAAGGACGC TGTTATTGTG GATTCGGAAA TGGTTCCTGT GAACGCCAAT 7951 CCTACTCAAG AGATACCTGC AGCTTCCGAG ACTGTAGAGT CTTCACCTGT 8001 AGCTCCAGGG AATACAACAG ATACCATGCC TGCAGCTTCG GGAACTACAG

8051 ACACCACATC TGGGGTTTCA GAGGCTGCGG CGGCTGAGGC TGCCGTGGAT 8101 TCTACACCAG GGACAGAGGA GGAGCCGAGT TTTTCTCTGA GGTATGCGCT 8151 TGTAGTTCAA AATGTTCCCT ATCCAGAGCC GCCTAAAGAA CCTGAGGTGA 8201 TGTTTACAGA TGAAGAAAAA AGTCTGATTT TAGAAGCTAC TCGTGCGCGT 8251 CGTATGGAGT TGGACTTGTA TAATGGCTAT TTAGCTGATT ATGAACTTTC 8301 TAAGGATGAA ATACAGAAAC ACGTTCCTGA TTTACCTGAG AATTGGCGTA 8351 CGAATTGGCG TTGGTCGGAG AGGCTCTATA AATTTTTCTT TAAAACAAAG 8401 AAAGAAGGAT TAGAAGAAAT TTTCTTAAAC AAAGAGTTAG GGAATATGAT 8451 TCTTGCCCGA GGGCTGGCGG CAACTCAGTC ACAAGCACGT ATTAAAGTAT 8501 TCAATTCTTT AGTGGCATGG CTCTTGCAAA GCTTTAACGT AGGGAGGAGC 8551 TGTACAGCTA AACCTCTTCC TACGTCAAAA CTAGACCTCT TTAAATCGGA 8601 ATTCGAGTCT AAGCCTAAAA ATAACATCTT AACGGAATTT TTGGTGGCCT 8651 CTGATGAGGA GATTCTCTTT AAGGGGCTAC GGGTCCTAGA GCCTGGAATC 8701 GAAGGTTGGT ATGACCATCC TGATCAAGCT GGAGAGATTC GGTCGGTACT 8751 CGAGGGTCTG GTGCAGGCTG GACGTATTTC TGGATATTGG GAGAATCAGC 8801 CGTTTGGGAG ATTTGTCCTT AGAGGAGTTG GTGAAAGACG TACCGAGCTT 8851 GTAGAGCTTT TGGAGAGTTT AGTTGCTTCT GGTGAGATTA TGCAGTTCTT 8901 TGAGTCTTCG GATGAAGAGG GTGCTTTTAT TATCGATAAC GAACCTAGCA 8951 AGACTGCTAT GCTAAAACAG CGATTTAAGA GTTGTGTCAG GACGAAGCTT 9001 GTCGGGAGTT TTGCTGATGA GAGTCTTCCC AGAGGTAGGT TTACCATTTT 9051 AGTTTAGCGT GGGGTAGAGC ACTCCACGAA TCTTAGGGAG CTCCTTGCGA 9101 CCAAGCTTGG AGATCCTCCA TGTTTTATTG TTTCTCTAGT AGCCAAATCG 9151 TAGCCGCTCC TAGGAACAAT TTTTTCTTTT TCGCAATATA AAATCCTGAT 9201 TTAGAGAATA GGTCTTCAAG ATCGTGGTCC TTTGGAAGTT GCTGGATACT 9251 TTTGCTGAGA TAGCTATAGG CGTCGGGATC TTTAGAAACA GACTTTCCAA 9301 TCCAGGGGAC GACAGCACGC AAATAGAGCT TATGGGCACT ATAGGTAGGG 9351 TGTGTTTTT TTGGAGGTGT GAGCTCTAGA ATGCCCAGTT TTCCAGAAGG

9401 CATAAGCACT CGGGAGATTT CTTGTAGGGC TTTATGTGGA TCCGAGAGGT 9451 TCCTGAGGCC ATAGGCCATC GCTGCTAGGG GATAAGAATG ATTCTCCAAG 9501 GGCAGTTGAT TAATATCGCT ATGAATAAAA GAGCAAGAGC CCTGGGGAAG 9551 GTGTTGTTTT GCAATGTCGA GCATTGCTGA GGAAAAGTCG ACGAGAGTTA 9601 CTGATGCTTG AGGGTGTGCG GCAATATAAC GCTTCGCGAC TTTTCCTGTT 9651 CCTGCGCAGA GATCCAGGAG AGAGTATCCC GACCCTAGGA TCTGGATCAA 9701 AGAGCGATTC CAGAAATGGT GCATTCCTAA AGAGAGTATT GTATTTGTGC 9751 GATCATACTT ACTCGCTATG GAATCGAAGA TCTTTTTACA GTCGGGCTTG 9801 TTGGTAGAGG GTTCCATAAT ATTCCCGGAA TTTTTCAAAG CTTTCGTAGT 9851 GTTCTTCTC TAGACGGTAC TGGCATAGGG CATAGTATTC TTGAAGAAGA 9901 GAAGGGGCA GACCTGTATG TTGATGAGCT TCTTTAAGGA CTTCTTCGGG 9951 TGAAGATTCG AACTGTTGGA GGGCTTCTTC CATCGCAAGG TTGGGTAGGG 10001 GATGTTCTTT CCAAGAGGTG CTGTGTAGAA GAAGAGCAAA TACAAAAGGT 10051 AGCTTTGTAA GATCATACCA CCCCGAGGCA AGGTCATAGG TTACAAATCC 10101 AGGAAGTACA GGATGTTGTA GCGCTGCATC TCCGATTAGG AGGAGGCCAT 10151 CATAATTTC AGGGGTTTGT CTGAGTACTT TTGTAGTTAT GAATCTTAGG 10201 ATATGAGGAG TTGGGATGCG CCAGAGATGA CGACAAAGCA CTTTTAAGAG 10251 TCCTATAGAG GAGCGACTTT CTAAAGTTGC GGCAATCCGA GGTTGCGGTG 10301 AGTTAAAGAA AGTGGGAGCT GCATAGAGGT TTACACTGAG GATACGTTGG 10351 TTTGCTGCAA TTCCAAAGCC GGGGACATAC CCCAAGTTAT GAGAGATAGC 10401 TCCTAGGGAT GAGGTCAAAG CAACATCGAG TTTCCCTTCG ATTAGCAAGT 10451 TGAGGAGGTC TGCAGGGGGA GCAAGAACAC AGCGAATATC GTTTCTTTTT 10501 ATGAGTTGTA GGGACAGCGG AAAGGAATTA ATATAACTTA CGCAGCCTAA 10551 GCTTATACAT GGCTGGAGTT GGTTAGACAT GGCGTTCTCC CTTGTTGTGT 10601 GATGAGGCC GCCATTCCCT CAGCGTCCAT TTTAATAGGT TCTTTAGATG 10651 AGGCCATCTG GAAAACCTTT TCCCCCATAT GTGTTGAAGA AAGGTCATTA 10701 GCACCACAGG AAAGGAGGTC TAGAGCTGCC TCAATACCTA GGTAATTCCA

10751 TAAGGCTTTC ATATTGGAAA AGTTGTCTAA GAAGATTCGG GCTACTGCCA 10801 TTAAAGATTT TAGAGGGATG GCATGACCCT GGCCTGATTT TCTTAATCTT 10851 TTTCCTAGGA CATTATTTTC TTGGGCGAAT TTTAGAAGTA TGAAGTTTTT 10901 AAAGCCCTGA GTTTCGTCTT GTAAGTCGCG GACTTTTACC ATGTGGGTGA 10951 CGAGGTCTTC AGGTCCTTCT TTATGATAGC AGAGCATGGT TATATTGCTA 11001 TGGATTCCCA GTTGATGAGC CATCTTATGG ATGTTGAGAA AATCAGAAGA 11051 AGAAAGGCGT TTGGGAGCTA AGAAATTACG TATTTTGTCG ACGAGGATTT 11101 CAGCTCCTCC TCCGGGGATG GAATCAAGAC CCGCATCTTT TAATGTGAGA 11151 AGAACATCGC GAATAGAAAG GTTATCAAGA TCTGAGAGAT AGGCATATTC 11201 AATGGCAGTA AGAGCTTTGA TATGGATCTG AGGATCGTAC TCTTTGATTT 11251 TAGTAAATAG ATCGGAATAG TATTGCAGAT TGCAGGAGGG GAAACAGCCT 11301 CCCACGATAT GTACTTCTGT AATTGGAGTT TTTATATTTT GGATTTGCTG 11351 TAGAAGATCA TCTGGGGAGT AGAGCCATCC TTTAGGGTCT CCAGGTTTTG 11401 CATAGAAAGA GCAAAATTTG CAGCTGAAGT CACAGAAATT TGTAGGATAG 11451 AGGTACAAGG TTGAGGAGTA GTATACAGTG TCGCCAACCC GTTGTTTGCG 11501 AACTTGGTCT GCAAAATTCC AGAGTGTGCG TTGATCTTCT TTATTCGTGA 11551 GGAGGAGGAG ATGAAGAGCG TCTTCACTGC TTAATCGTTC TTGGGCATCC 11601 AGTTTTCGA ATATGGAGTA GAGGGGGGAA GTTTTAGGGG GCTGTGGGAG 11651 GCACGTCGTC ATTTGATGAA CACTTTGATG TACTATTCTC TCGAGATTTT 11701 GTAGCACAGT GCTCTGTTTT GTCACATGTT TTTTTTTGGC AGCAATCTGG 11751 TTTTTGACAC CCTTTAGAGA GGGGCCTGCC AAGCAAATGG GAACCTACAA 11801 GTAGAATACC CATCCCTAGA CCTAACAATA CTGTGGCACA GCAGATTACG 11851 AGAAAAAGTG TCATCATAAG AAATCCTTAG ATAGGATAGT TTCTTAATTT 11901 AAATCCACCC AGATTGGGGA ACTCCAGGCC ATAGCATTGT CTGCCTGAGT 11951 GACCCTGAGA TAGTAGAATA CAAAAGGTGC TTTACCGTTT GGATCTTTTA 12001 GGGTCACTGA ACTTAGGGGT ACCATATCAT CGTATTCATA GTCCAGGTTA 12051 TTGCTATCGG GGAAGAAGGT ATGGAGAACT TCGCCATTGC GGATGATTTC

12101 TACAGTCTTG AGTAGGGCAG TGCCTGCCAC ATGACCAGAG ATGTGACGGT TGACGTTGAG TCCAGGTTTC GACCCTGTGG AGAGTTCGGA GCCCATAGGG 12151 12201 GCTGAAGTGA TGTTGAAGCT TAAGACGATC CTAGGTCCTG TTGTAGCGTA 12251 GCAATGACGT GCGAATAAAG CTTCAACAAG AGACTCTCGG GTATATTTAT 12301 TACAAATGAT AGCCGTCAAC CCTGGGGAAT ATTGCACTTG CGGAGAGTCA 12351 AAGTAGTCTT TATAAATTCC TCGATCGTCG AGACCCCCAG CAACAAATCC 12401 GAAGCGGAGA TTCTTCTTTA ATCCTTCAAT TACTGTACCT CGAGGATCTT 12451 CGCTATCTTT ACCTTGGATA GGGAAGGGGT TGTTTAGAGC GGCTGTGGTT 12501 TCTGAAGATC CCCAGGCATT ATAAATTTCT ACAACTCTTT CGAACTCGGG 12551 GTAGAAATTC TCAAAGTCAA AACCATGTTC TTTAGAAGCT GTGAACGAAG 12601 GAATAGAAAT CATGTCGTGG TTGACAGTGC TTTTATAGAG CTTGGCGAGG 12651 GGAATATGTT TGTATTCTTT GTGTTTCGAG TGGGACTTTG TTTCCTTGGT 12701 ATGAAGGATG TGACGCACTC CCTCGAGATG AGGTTCTCCG CTATATTGGA 12751 ATCCGGATAG TGTGATGAAG CGATCTTCTT CATTAAAGTC GGAGACAGTT 12801 TGATTGATGA GCTTCCAAAT ATCTGGAGAG AGGTTCTCTT GATTTTCGAA 12851 TGATGAAGAA GCATAGAAAT TCAGAGCGCG GTCATCTCGG AAATAACGCA 12901 TACAAGTTTC AATATTTTCT TCAGAGTCGA CGCGTTCGGA TTCGCCGTGG 12951 AGGAGACCCC ACATAAGATT CGGGGCGGAG TCAGCGAAAC ATTTGATAGG 13001 GGCAGAGATG AAAATTTCTT GTGTAGAGAG GTTTTTCAAT TGGATGCGAT 13051 AAATTCCAGG CTCATTGAAA TAGAGATTAG GAAGAATAAC AAAGCCTGTT 13101 TCTGGGATGA AGAGCTGCCA ATTTAAATTT TCTCTAAGAT GCTCGTAGGA 13151 AAGCTCGATT CGGGTCTCTT CAGGAGAGAA GTTGGTGAGG TTCCCGAATT 13201 CGTCTTCAAA TCGCACGGTG ATATCGAAGC GTTTGTTTTT AACGACATAG 13251 GAGGGAGTAA AGATCTCTAT TTTTTTTAGG ACGTTTCCGC GGATATCCAT 13301 AGAGAAGACA TCGGGTTCAT CATAGTTTCC TTCTCCTGTA GGATCGATGT 13351 AGAGGTAAAA GGGTTTGCGA CGTTGTGCGA AAAGTTGGGC TCCGTTCCCA 13401 GCATCATCGA CTTGAGGATG GTTTGGAGAG GCTCCCATGA CAATAGTGAG

13451 GGTTTCTCCT ACTTGAAGTT CGTAGGGGAG AGTAAACTCG AATTGTGGAA 13501 CGGGATTGTC TTTTACAGGA ATGGCGGTTG CTTCGATGAT TTCGCCTTCT 13551 GGCATTTCTG CGTAGATTAC GTTTCTAGTT TGGGAGAGAT CTGTCGCGGG 13601 GGCTTCCCAA TCTGTGGGTT TCCCACTTCC TGCTAAGTCA AATTTACATT 13651 TGGTTCCAGC TGGTAGTGGT GTGGCAAGGG AATAAAGAAA TTTCCAAGTA 13701 GAAATTTGCC CTGCTCGAGC TATCGAAGGG TTAACGTAAC AAACAGATCG 13751 TCGCATAGTA AGAGGGAGGC TTTATATGAC TTAAAAGCGC CATCATATAC 13801 TAACAGTGAG GTTTTTCTCA ATCCCCGTCT TTGTTTAGTG TTTGTATCGC 13851 TTCATCCACA GTATTGAATA TTTTAAAGTA AGAAAGGAAT CCTGTAACAT 13901 AGAGAGTTTG TTCTATGGTT TTTGGGACTG TAGTCAGGAC AATTTTCCCA 13951 GAATGTTGTC CTACTTGATG GTAGCTTTGC AGTAGGACTC GGATACCTGC 14001 ACTGGACATG TAATCGAGGT GAGCACAGTC GAGAATGATA TTTTTGGATC 14051 CAGCTGCTAG GGATTGGGAA ATATTTTCTT GTACTTCTGG AGAAGAAATT 14101 CCATCAAGTT TTCCGTGGAG ATGAAAGATT GTTGTTGAGC CGTGTTCTTC 14151 TTTTTGGATA TCACTCATCT AGATAGTTCT CCTAACTATA CGGGAGCTTA 14201 AGTTTTCACT CTGATAAATC TTTAGCTTTT TTGCAAAGAG ATTTTTATTG 14251 GTGATGTTTG AGAATTTCGA TTGGGGGGAG GCAGGATGGG ATCGTGGAGT 14301 GCAAGGAAAT CGGTCCTAGG ATGTTTACAT TCTTAGGAGA TATTGAATTT 14351 ACGTTTTCTT GGTGTGATTT TTAGTTTTCC GACATTTCGT TCTGTGCAGG 14401 TAATGATTTC TATATCGAAG TTTTCGTGAT GGATACGCAT TCCTTTTTGG 14451 GGAACAGCAC CCACTTTATG GAAGACATGT CCTCCTAGTG TATCGTAGCT 14501 ATTTTCATGA TCGATTTTCA AATTGAAGTA CTCTTCAGCG TCGGAGATAT 14551 TCATTCTTCC ATCTACAATC CAAGAGCTTC CGATTTTCTT ATAAGGAGTA 14601 TTTTCTTGTA CGTCGTGCTC GTCTGCGATC TCTCCTATAA TTTCTTCGAT 14651 AATATCTTCC ATGGTAGCGA TGCCTTCTGT GAATCCGTAT TCATTGACTA 14701 TGATGGCTAG ATGGCGATGT TTTTGTCGGA ACTCTTGGAG AAGAGAGGAG 14751 GCTTTTTTA TTTCTGGGGC ATAGAATGGG GGTTTGCTAC TGAGGATATG

14801 GGTTGGCTGA GGTCGTGGCT GCTTGTATAG AGCAGTAAGA GATCTTTAAC 14851 AAGAAGGATT CCTGTGATGT TGTCTAAGTT TTTTTTATAA ACGGGAACGC 14901 GACTGTAGCC TTCTTCGCTT ACGAGAACCA GAGCTTCTTG TAGTGTAGTT 14951 TCTTCGGGAA GTGCGAAAAT ATCTACTTTT GGGATCATGA CTTCACGGAC 15001 AATGAGGTTA TCAAAAGCGG AGAGGGCTTC GGAGAGCTGG CTTTGAAATG 15051 ATGTTGAAGA TCGTACTTGT TGGTTAGGGC GGCGTCTGTA AAAGAGCAGT 15101 TGCAGTGGGA AGAGACCGAG TTGGAATACC GAAGCTAGAA AACGGAGGTG 15151 GGCGGTGGTT TCTTTAGGGA CTTTTGTAGA GATCCATGGG GGGAGGAATC 15201 CGTAAGCTAT CAGGGCGCTT AGAGAGTATA GGGGCCAGAA TAGGAGATCT 15251 TTGTGAGCTG TTTTTGGAGG GAGGAGGGTA TAGAGTTTTG TCCCGAGAGC 15301 TCCATAGAGG ATGCAGAGCA GCGTGGCGAG AATTGTAGGA GCACTGGGGA 15351 AGGGGGGATA CTCTCTTCCT TTATCTTTGA AGAAGCGTTG GTTTAGGGTT 15401 TTTAGGAATT TTGAGGATCC GTGACAGGAC GGTTGCGTAA GCCCGAAGGC 15451 TAGGAATAGA AGAATACAGA ATATGGCTAA AAGAATATGG AGCATGTTAA 15501 GCTGTTAGCA AAGCATGTTT TTTTCTTAAC ATACACAGGA TTTGATTTTC 15551 TTTAACTCTC ATTTTTCTCT TTTCTTCTGA TGAGGTGTCG TCGTATCCGA 15601 GCATATGGAG AATAGAGTGG ACGAGGTATC TCGAGATTTC TTCGTAGATA 15651 TCCTCTTGGT TTGGGGATGT GTTCTCTAAA AACCTAAGAG CGGCCTGTGG 15701 GCTAATGAAT GCTTCTCCTA AAACATGAGG ATAAGCGGGA TCTCCGGGAG 15751 CATCAATAGG CAGAGTGATC GTATCTGTTA GAGAAGGATC AGCAAATACC 15801 TTATCATGGA GTTCTGCAAG AGCTTTATCT TCTAGGAAGT AGATAAAAAT 15851 TTCATTAGTT GTTACTTTTA AGTGCTCTAA GAGCGTAAGA ACCAGCTTCT 15901 CTACAGAAAC CAAATGAATA GGAATACATG TTTGCTCATT GGAAACATGT 15951 ATTTTGATCT TTTCTTGCGT CACGCGAATG AAATTCCCAT AGAGAAAAAC 16001 AAACTTATTT TAAAATAGGG GTCTTAGGTA AACCTGTGAC TTTTTTCGCT 16051 GTACTATCAT TCCAACGGCC CAACTTACGC AAGACTTCTA CTCGCTCAAA 16101 ACGCTTCAAA ACATTTCTTT TGGTAACCCC TTTGACAGAT TTACCATAAC

16151 TACGATGTCG AGACATAATC CTGCTCTAGA AATAAACCTA TTTTCGGGAT 16201 AAAAATGCTA TCACTGGTGC TCAATGCATT CGTATGCAAT TTATATACAA 16251 TTCTTGGAGC TGGCGCTGGA TGCACAATGG CATACTTAGG TTTACGTTTT 16301 TTAGGACTTT TCGCTCTGCG CCGAGCTTGT TTACTCATTC GTGTCATAAA 16351 TAACTCGATA TTGAATTTTT TAATTTCTCC AGACAACGGA AATTGAGGAT 16401 ACGGAGTACT TTATAAGAAA AAGGATAGTA AAAGAAGAGT TTTTTTCAA 16451 GAAGATGACG TCTTTTAGCT GCCTTGATCT TGGTGTAGCT CGTCAGGGAG 16501 GGGAGACGAG GGGCATGCCA TGGGGGTTGA GAGGACTCCA CAGGAGATAA 16551 TATATTIGAA AGCATCTTCG ATTTTCATAT CTAGGAATAC GATATCAGAT 16601 TTTCTAAATA GGGTAAGAAA CCCTGAGGTG GGGTTGGGTG TTGTTGGGAT 16651 GAAGACCGTG ACGAGGGGGT CGTCTTCCTT TTCTCCTGTG CAGCATACTG 16701 TGGGTGCGTC TCCAGCGACG AGACCGATGC ATTGAACATT TGCGTTAGGG 16751 AAAGGAACCA TAACTACTTG TTTGAAGGAT CCTGATTTTG ATCCAAATAT 16801 GGTAGTCATG ACTTGTTGCG CAGCTTTATA CACTGTTTTA ATGATGGGAA 16851 TTCGGTGTAA GATTTTGTCG TAGATAGAGA GTAGGGATTT AAAAATCATA 16901 ATTCTCGTGA GGAAACCTAG GAGCACTGTG GCGAAAAAGA GACCGAAGAG 16951 TAAAATGATT TGCAATACGA ATTTTAGAAG AGCTCTATGT TTAGTATAAA 17051 AGGAAGTTCA TGATCATAGT AACAATAGCA ATAGTAATTG CTAGAGGAAG 17101 GAGAATAACA AGTCCTGTAA TAAAGTATTT TTTCATGATT CTCCTGCAAG 17151 ATATGAGGAA ATGGGCATTT GTTTCTTTAC TATACAGCTT AAGATTATTT 17201 AAGATAAAAC TTTTCCCGAA TCTTCTGGGG ATAGGAGAAA TCTCCATGGG 17251 ACATCACGAT ACTCTTGAGC ATAATCGATG CCGATCCGGG CAGTTGCTGT 17301 TAGAGTCCCA GAGATTTTTT CTTTGCTGAT ATAGAGAGCT GGGGTATTTA 17351 GGCGTTGCCT ATTGTTTTCC AAAGAGATTC CTAGAGCTTG GCACACTTTT 17401 CCGGGTCCAT TGGTGAGAAG GTGTGGGGGT TTATCTCTCC ATTGGCGGCG 17451 TTGGATCATA AGTTCTTTGC CTTGATCAGG AAGGATGGCC CGGATCAGGA

17501 CGGCATGGGG AATGTCCTCA GGTCCAGTGA CAACATTCAA TAGGTGATGC 17551 ATGCCATAGC AACGGTAGAG GTAAGCAGAG CCTCCTTTCA GGTACATCGC 17601 TCTGTTCCTC TGAGTTTTTC TGTAGTTGTA GGCGTGGCAT GCTTTGTCAT 17651 CAGGGCCACG ATACGCTTCG GTTTCTACAA TGTAACCTGA AGTTATCAGA 17701 CCCTCATGTG TTGTGATGAG TTTATGTCCT AAAAGCTGTT GCGCTAGTGT 17751 AATTACATCT TCCGATAGAA AAAAATGTTC TTGTAGCACG TTACGAGGCT 17801 CTTTTTTCG TTCCTTTTT CTTAGAAGGC GTTTTCTTTA TTTTCTTAGG 17851 TTTATCTTCT GTGGTTGTCG CTATAGACCA GACGATTTTT TGCGTAAGGA 17901 GATTCACGGA ATCAATAGTG ACTTTTATAG AAGCTCCAGG TTTCATTTTA 17951 TCTGGGATAG ATTCTGGAAG AGCGTTTTTC TTTAGGGAAT ATTCTTTAGG 18001 GAGTTCTGCT GCTGCAATGA ACCCTTCATG GCAGAATTCG GTCACTACAA 18051 ATGAGAGTCC TTCATGATTT GCAGTGATGA TATACGCATG GTATGTAGTT 18101 TTAGGTTGCT CTTGCAAAAA TTTATTTATG AACCGAGTTT TTTTGAGGTT 18151 TTCGAAAGAA TTTTCTGCTT TTGCGGATAC TCGTTCTTTT GTAGAGCATG 18201 CTCTTACGAT AATTTCGAGG TGCGTTTGGT CTATAGATAG GGGGTTGAAG 18251 AGAAGCCTGT GAACAATAAG ATCGATATAT CTACGTATGG GACTCGTAAA 18301 GTGGGTGTAG TAGTCGAGCT TAAGTCCGTA ATGACCTTTA TTTTCTGTAG 18351 AGTAGGAGGC TGTTTTCATA CTTCGGACAA ACTGCGAGTG TAGAACTTGC 18401 TCTAGGGGAT GTCCTGCTGA CGTAGTTTGC AAAAGGTATT GGTAATCAGG 18451 TTCTTGTGTG GGAGTGAACG TGATATCAAA GCCCATGTTT TTTGCCAATT 18501 CTTGGAAGGC GAGTAGGTTT TCATCATTGG GAGGTTCGTG ACTACGAAAA 18551 GGTAGAGAAA CGCCTTGATG GGAGATATGA TAGGCGACCA CTTCGTTTGC 18601 TTTAAGCATA AACTCTTCGA TGAGTTTATG GGAGAAGGTC TGGTGGTTTT 18651 CTATCAGAGC TACGGGTTCT TGAAGATTAT CCAAGGACAT AGTGACTGAG 18701 GGGAGGACAA AGCGAATGCA ACCACGTTCT TCACGGATAT CGGAAAACTT 18751 TTTACTTAGA GTGGCCATCT CATTGAGGAT TTTTGAGAGG GGGTGGGAGT 18801 GTTTCTTTC AATGATGTTA TCGACTTCAT CGTAGGTCAT ACGATATTTG

18851 CTTCGAATGA CGCTACGGAA AATCTGGTAA TCTGAAAGAT GACCTGATTT 18901 TGTAAACGTC ATAAATACGG ATACAGCGAG TCTATCAACG TTTGGTTTTA 18951 AGCTGCAGAG ATTATCAGAG AGTGCTGATG GCAACATGGG AATGACTTTC 19001 CCTGGGAAAT ATGTAGAGTT ACAGCGTTTA GCAGCTTCTT TGTCTAGGTG 19051 AGAATGTGGG GTAACGTAGT GGGAGACGTC TGCGATGTGT ACACCAAGAA 19101 TGTAATTGTT ATTATGATCG TAGGTGAGGG AGATGGCATC GTCGAAGTCT 19151 CTGGCTGTGG AAGAGTCTAT GGTGAAACAG AGGAGATCAC GGAGATCTTT 19201 GCGAGAGTGG AGAACTTGGG TAATGTGTTT TTGAGAGAAA AGGCTTGCTT 19251 CTTCAATGAC CTCTGGGGGG AATTCTTCGG CAAGGTTATA TTCGGCTTGA 19301 ATTGCCTGAA AGTCCGCTTT AGCGTTGGTG ATGTGGCCAA TAAATTCGAG 19351 CATTTGTAAG GCTGGAGAGG CTCCTTCTTG GGGTTTATCT ACCCAGGGAG 19401 GAGTGCTCAG AAGAATGCGA TCGCCGATTT TGTAAGTGCG TCCGGGAAGG 19451 AGTTCTACTG GAATTAAAGA TTGGGATCCC GACATGCTTG TGTAGGCAAG 19501 TGCTGATGTG GGACTGACTA GTGAGGTGAT CGTTCCTACG AGTGTTGTTT 19551 TTCCTCTTGC GAGTACTTCG CTGATAGTGC CTTTGAGTTT TTGTCCGTCT 19601 CTTGGATAGG GAAGCACGGA GACAATCACG TGGTCACCAT CTAGAGCCCC 19651 GCGTAAATCT CGGGCGGGAA CAAAAATATC AAATGGGTAT TCTTCGGGGT 19701 TGTCGGGAGA AACAAAACCG AAACCTTTTC TAGCATGAAC AAATAGGGTT 19751 CCTGGAATAA AAATCTTCAA GGATTTACCG TATGTTCTTC TCCCTGGTTT 19801 TCTTTTTGGT TTTTTCAACA ATTGGGCTCC GCCTGTAAGT TTAGGACATT 19851 GAACGAGAA TCCCGTAGTT TCACTCGTGA ACTGAGTGGC TCAACAAAT 19901 TTTCCCTTTT AAGTGGGTTT TGTGATTATG AAAGGAGCAG TCTCAAATTC 19951 AAAGCCAATT GTACTAAAGA AGCTCTGTTA TTGCAACTCC TTGATCAGAG 20001 AATAAGAGAA TGGAACTGTT TTCTGTTTAT AAGGAAGTTT CCCATCCTCT 20051 TATGAGGATG GGAATAGAGA AACTTAAATT GAAAATTTTG ATTACTTATC 20101 GTCGTTATCA ATAATTTCTA CATCAGCTTC TTCGATATGG TCTTCTGAAG 20151 AACCGTTATT TGAAGGAGGC TTCGTACTGA AACTATGTTT TTTCAAATCT

20201 TCTGTATTGA TGTTAGGTCC ACCTTTAGCA TTGGCTGCCG ATGATGCTGC 20251 TGCTGATGCA GACTGCGATT GCATAGACTC TCCAATTTTT TGCATATGCT 20301 TGCTTAGGTC TTCAGTAACC TCTTTAATTT TTTCAATAGG AGCGTCATCT 20351 TTGAGTGCGT TGCGCACGTT TTCGATTCGC TCTTCGATTT CTTTAACTAA 20401 AGTTTCAGGA ATTTGCTCCT TATAATCTTT AATAGCTTTT TCGGCTCTGA 20451 AGATCATGCT ATCGGCTTCA TTTTTAGCAT CTGAAGCTTC ACGACGTTTT 20501 TTATCTTCTT CCTTATTAAT TTCGGCATCT CGAACCATTC TTTGGATTTC 20551 ATCTTCTTGA AGTCCTGAGC TTGCTTCGAT ACGAATTTTC TGTTCTTTAC 20601 CGCTGGCAAC ATCTTTAGCT GAGACATGGA AAATTCCGTT TGCATCGATA 20651 TCGAAGGAGA CTTCGATTTG AGGATGGCCT CGAGGAGCCG GAGGGATATC 20701 TGTAAGATCG AATCTTCCGA TTTCCTTGTT ATCTTTGGCC ATGGGACGCT 20751 CTCCTTGGAG AACTACGATG GTAACCGCAG CTGGTTATCA GCAGCTGTGG 20801 AGAAGATTTG TTTTTTCTGT GTAGGGATTG TAGTATTTCT CTCTACCAGA 20851 GTCGTCATGA CGCCTCCTAG AGTTTCGATA CCCAGAGATA GGGGGATAAC 20901 GTCTAGAAGT AGAACATCCT TAACTTCTCC GCCAAGAACA CCACCTTGAA 20951 TTGCGGCTCC AATAGCAACA ACTTCGTCGG GGTTGACTCC TTTATTAGGC 21001 TCTTTGCCGA AGAGTTCTTT TACAGTTTCT TGCACTGCGG GCATTCTTGA 21051 CATACCTCCA ACTAAGAGAA CATCATCGAT ATCCTTAGCG GAAAGTTTTG 21101 CGTCACTGAG TGCTTTGATG CATGGAGATT TTGTTCTTTC GATTAGAGAG 21151 GCTGCGAGTT TCTCGAATTG CGCACGTGTG AGTGTCAATG CAAGGTGTTT 21201 AGGTCCTTGT GCATCCATTG TGATGAATGG CTGATTGATT TCTGTGGAAG 21251 AGACTCCTGA AAGTTCTATT TTTGCTTTCT CAGCAGCATC TTTAAGTCTT 21301 TGTAAGGCCA TATTATCTTT GCTAAGATCA ATGCCTTCTT GTTTTTTGAA 21351 TTCTTCGATC ATCCATTGA TAATGACTTC ATCAAAGTCG TCTCCACCGA 21401 GGAGAGTATC TCCATTTGTA GATAGAACTT CGAAGACGCC ATCACCGATT 21451 TCTAGGATGG AGATATCAAA AGTTCCTCCA CCAAGGTCGA AGACAGCGAT 21501 TTTTTTATCA CCGACTTTAT CGATTCCGTA GGCAAGAGCT GCTGCGGTAG

21551 GTTCTGGAAT GATACGTTTT ACATCTAGAC CTGCAATGCG TCCAGCATCT 21601 TTTGTGGATG CTCGTTGAGA ATCATTGAAG TATGCGGGGA CGGTGATCAC 21651 TGCTTCTGTG ACAGTTTCGC CTAGATAAGC ATCAGCTGTC TCTTTCATTT 21701 TCATTAAGAT TTGTGCGCCA ATTTCTTCTG GAGTGTATTG TTTGCCATCA 21751 ACTTCGAAAA CGGCATCACC TTTAGATCCG GAGGTGACTG TATAAGGAAC 21801 GGTTTGGATT TCCGAAGCTA CTTCAGAGTA CTTACGGCCA ATAAAGCGTT 21851 TTGTAGAGCC GAGAGTTTTT TCTGGATTTG TCACTGCTTG ACGTTTTGCT 21901 GGAATCCCCA CTAATTTCTC ATTACCTTTG AAGGCAACGA TCGATGGCGT 21951 GGTTCTTGTT CCTTCGGATG ATGTAATTAC TTTAGCTTGT CCTCCTTCCA 22001 TAACAGATAC GCAGGAGTTT GTTGTGCCTA AGTCTATACC TATAATTTTG 22051 CTTGATTTTT TGTGTTCACT CATGTTTGGT ACCTAATCTC TAGGGGTTAT 22101 TTCTATTCTT TATTTTCTTT GGGAGTAGGA GCTTTAGCGA CTTTAACTTT 22151 AGCTACCCGA ATCGGGCGTT CTCCTATTTT ATATCCCTTT GCAAACTCTT 22201 CTAAAATCGT CCCCTCAGGA ACTTCAGAAG TCTCTTCTGT TTGCACCGCT 22251 TCGTGTAGGA AGGGGTTAAA CTTTTGGCCT ATTGAAGAAT ATTCAATAAT 22301 ACCTTTTCC TCGAAGATTT GTTTGAATTG GTTGAGAATC ATGTTGAATC 22351 CGAGGGCCCA ATTTTTTACA TCGTCGGACA TTTGTGTAGC AAATCCGAGG 22401 GCTTTCTCCA TGCTTTCTAT GGGATTGAGA AAGTCTATTA AAGTATTTTC 22451 TAAAGCATAC TGCATAAGTT CTTGGCGTTC TTTTTGTAAG CGTTTTCTAG 22501 AATTCTCAGA TTCTGCTAGA GCCATGAGAT ACTTATCGTT TTTTTCTTTT 22551 AATTCGGTTT TTAGGGTGAC GATTTCCTGT TGCAAATGTT CAACTTCATT 22601 TTCGTTTTGA ACATTGCTTT CGTGTTGTTC CTCATTTTCA GGTGGGGTAT 22651 CTGTCATAAC GTCTCCTTAG AGGGTAATAG TTTTATAGAA GAGTACTCCG 22701 TTCTTAAAAT AGGCTCATTC GAAAGCTTAC AGTTAGAGGT GAGTGGTCTT 22751 CTGAAGGATA GTTTAAATTT GTAGAAACTT TGTGTCAGGG TTTCATTTAT 22801 TTTATTCGCA AATAGTTTGA GCAAAGGAAG AGCTTCCTTA TAAGGAAGAT 22851 TGATCGGGCC TAGGATACCT AAAGCTCCGA GTGGAGAGCG ATTCATATAA

TAGGGAATAG TAATTACAGA ACATCCTGGA TTCGAGGTCC CTAAAATATC 22951 AGAAAGCTCC TTCCCTATGA ACGCTGTAGC TCTTCCTTTA TGCATTCCTA 23001 TATTTAGAAG CTCACACATT TGTCTGCGAT TTTCAAAAAG AGAGAGTCCT 23051 AGAGCTAGAA CTTCAGGATC TTTAAACGCT TCGTATTTCA GTAGTTTCGA 23101 CATTCCTGTT TGATAGAGAT CTTCTTCACT AAAGTTGCAG TAGCGTGTTA 23151 GATAGCGGAC AACCACCTCA TTATAGAGGG ACATGCTCAG GTGTTCTTCT 23201 TTTTTCGAAA GTTCCTCATT TGTGGGGAGC TTTCGGATGT AGTTCTGCAG 23251 GAATTTTCT ATACGTTTGA TAGAAAGAGT ATCGCAAGCT TCAGGCAGCC 23301 ATAGGGTGTC TGTGAAGATC TGACCAAACT CCGTAGAGAG GATGGTGACA 23351 GCTCTTTGCT TATCGACCTG TGTAATTTGA ATATTGGTTA CGGAATCATT 23401 TTCAAAGCGT GGGGAAGAAA AAAACGTAGG CAGGTCTAGG ATTTCTCCAA 23451 GAAGTTCCGT AGCTTTTTGT AGATCCTTGA TAATATTGCG ACTTTCGCTA 23501 GGAAGCTGAC TGATCTTATC AAAAATGGGG GCAGAAATCT CAGCTTCTGG 23551 GCATTCTTCT TGGTGATCTA CATAGTGACG TAATGCTAGG TCTGTAGGGA 23601 TTCTTCCTCC GGAAGTATGA TTTTTTTTTA AGAATCCTTC AGCTTCAAGT 23651 TCTGCAAAGT AATTTCTTAT AGTTGCCGTA CTCAAATCAG AGCAAAAACT 23701 TTCCTTTAAA GTTTTAGACC CTACAGGCTG CCCTGTTTTT AGGTACAACT 23751 CTGTTGTAGC AAACAGGATA TCAAGGATTT TTGAATCTCG CTTTGAGACT 23801 TTGGATCTAG CCATCTCGAG TCCTACTAGA ACAATCGTAA CTGAGAGCAT 23851 TATAGGAAAA GAACCCGAGA AGGTCAAGAG AATTTAGCAC TCGAATTGAA 23901 TGAATGCTAA CTTTTTTACG AGGAGGGCAG CGATCAAAGA GCTAGGCTAA 23951 GTGATTCTGA CACCAAGTAG GGAAGGCCTC CGGGGAGACT GTATACTTTT 24001 CTCCAGATCG GGATTCAATT TCGAATATTC CCGAAGATTG GTAGGACTTT 24051 CCTAAAATAA GCTTATAAGG AATGCCGATA AGGTCACTGT CTTTAAGTTT 24101 AAATCCGAGT CTTTCATCTC GATCATCAAG AAGGGGCTCA TAGCCTTGAC 24151 TTTGTAGCTC ATGATAAATA GTTTCCGCAA GCTCTTGAGA TACAGTGTCT 24201 CCTCCGTTAA AGGCGATAGT GATAGAGAAG GGAGCGAGTG CTTTTGGCCA

24251 AACAATACCA CGGTCGTCGG CAAGCTGTTC TACACAAGCG GCTAATGTTC 24301 TTCCGACTCC AATGCCGTAG GTCCCCATCC AGCACTGCTG GGTTTGCCCG 24351 TGTTCATCTT GGAAGTTTAC CTCAAAACTA TCGGTATAGC GTGTCCCGAG 24401 ATTGAAAATA TGAGCAACTT CTATGCCTTG ATAAATGCGG TAAGGATGGC 24451 CAGGATTTTC AGGACATGTG TCTCCCTCTT CAGCGAGTAG AAAGTCACCG 24501 TATTGGGGG GGAGGAGGTC GCGATCCCAG TTTACATTTA CGTAGTGCTT 24551 ATCTTTAGCA TTGCCCGCAC AAACAAAGTT CGTCATTGGG GACGTTGTTT 24601 CGTCTGCGAA AAAGTCTATG GGACAGTTTA GGGGACCGAT GAATCCTTTT 24651 TCTGTGCCTA GAACGCGTTC GATTTCTTCA TCAGAAGCTA GAGCAATATC 24701 ATCGGCATTC AGTTTGGAAG CGACCTTCAC TAGGTTGACT TGCCGATCTC 24751 CTCTCATTCC AATGGCAATG AATTTTTCTT CATTTGAGTA GGAGAGTTTT 24801 ACGACAAGGG TTTTTAAAAT TTTATGTAAG GGGATAGAGA AGAAGTTTGC 24851 TAGAGCTTCT ATTGTTGTAA TCCCAGGGGT GGCCACTTCT TCGACGGGAA 24901 GAAACTCGCG ATCGTAGGCA TGCTGTGGAG GAATGGAGAC AGCAGCCTCA 24951 ATATTAGCTC CATAGGAACC GCTGACGCAG ATCGTGTCCT CGCCTAGAGA 25001 GCAAAGGACC TGAAATTCCT CAGACTTTCC TTTGCCGATT TTCCCTCCAT 25051 CAGCTGTAAC GATGACATAG GCAAGACCGA GACGATCAAA GATCTTACTA 25101 TACGCAGAGC GGAGTTTTTC ATATTGCTCG TTCATTTGTT CGGGAGAGTC 25151 TGAGAAGGTA TAGCTGTCTT CCATAAGGAG CTCTCGAGAG CGAATGAGAC 25201 CGAATCGAGG GCGAATCTCG TCTCGGAATT TTGTAGCAAT TTGGTAAAGG 25251 TGGAGAGGAA GTTGTCTTTT TGAGGAGAGC CATTGTGCAA CAAAAGAGCA 25301 GATGACCTCT TCATGTGTAG GAGCTAGGCA ATGAGATTTT CCTTCGCGGT 25351 CTTTGAGAGT GTAGAGCAGT CCTTCCGAAG TAAATGCCTC CCATCTCCCT 25401 GTATGTTGCC AAAGTTCAGC ATTGTGGAGA AGTGGGAGTA GAAGTTCTTG 25451 ACCTCCAATC GCATTAAGTT CCTCTCTAAT GATGTTCATC ATCTTGGAGA 25501 CCACGCGCCA TAACAGGGGT GTATAGGTAT AGACTCCTTT ACTTACTTTA 25551 AATAGGTATC CTGCCTTTTC TAGGAGCTCG TTTGAGAGCA CAGCAGCGCT

25601 TTTATTTGCA TTTTTTGAAG TCTTATAAAA GAGTTGAGAC GTTTTCATAG 25651 AGTGGGGCTG TTATCTTCAG TGATTGATCG TCAAGAGTCT ATAGAAAAAA 25701 TTACAGGCTA TTGCGATTTT ATATCGAGAG TCTTTTAAGC TAGTAGATTT 25751 AGTTTAGCTA TGAAGGAGGT ATCTTAACAC AGGGATTTTT CTAGATCAAT 25801 GTTTTGCGTT GGGTTCTTAA CTTAACAATT GGGGTCAGGA TTATGTATTA 25851 TTTTAATTAT ATTTTTGTTT TTATTTAGAA ATTTAATAAT CTCAACTTAT 25901 AATTAAGTGT ATACTTATTA ATCTTTTATT TTTGTAATTG TAGTACTATG 25951 TCGTCAGTAA ATCAAAGCTC TGGAACCCCG AATCCAGAAG AGGTAACTTC 26001 TCCTGAATCT ACGGAAGAAA ACAAAAATGT TGTTTCTTCA GATGAGGCGC 26051 AAGCCACGCA TGCTGTGGCT CTTCCTATAG TCACTCAACT TTCTCTTCCT 26101 GAAGGTGTGG GGACCTCATC TGAAGAAACG GCGAGTAATC CGAGGGTAGA 26151 CGAGATTGTA GCTGAAGTTT CTTCGAGTCG GGCGGTTGCT GATCAGATCT 26201 CATCACTTGT AGAGCGTGTT GGAGAGCTTT TAGACGACCT TAAGGGTGCC 26251 CAGTCCCTTT TCACTAGCTT TCAGTCAGAG TTGAAAAACT GTCTTCCGGC 26301 ATGGAAATCT TCAACGAGAA GACTCGAAAC TCGAGGTGCT GGGGATAATG 26351 CGGATATAGC GAGGCTGGAA TTATTTCGTA GCGATTACGA GGCTGTCTTA 26401 GGCCATGCGA ACCAGTTTCA TGGGAAGGCT CATCTCATTT TAAGTAAGTT 26451 AACAGATGTA CATCACAAGC TACAGGGACT CAGTCGTGAA GATCTTTCCC 26501 TGGCGTTTGA CAATAATGAT AGGGTTCTTG AGCATCTGGG TTCGTTAGGG 26551 CTTGATGTAG ACGCTGAAGG TAATTGGTCT CTTTCTTGTG AGAGGGGGAT 26601 TCCGCGACTG GTGCTTACTG CTGACAGTAT GCTTGTCCAG ATCAAGAAAG 26651 TGAATCTACC TACTGTAGAA GAATTGCGGA CTCTTCAGGG AACAACGGAA 26701 TCTTCGTCTG ATCCTAGGGT TGAAGAAGT TTGTCTTGCT GTGAAAGATT 26751 GCTCAATGAA TTACGTCGTC TTTGGGCGAA TTTTGTAGGT TTTATTTCGA 26801 GTTGCTATGA CAACATCGTG TTTGTTTTGA TGTGGATAGT GAGACGGATT 26851 AACCTTTTGC CTGGGCTGGG GTGTTTGCCT TTCCATAATC CCGATGCTTC 26901 TCAAGAAGAC CAGAGGTCTT CTTCCGGAGA GCGTTCTACA AGGAGAGAAC

26951 GCCTTTCTCG GCGATCTGAC TTATCTGAGG AAGAGATGAT TGTGAGAGCT 27001 GAGGGAGAT CTATACATCC TGAATCTCCC CATGGAGATG GCCGTAACCA 27051 ACCTAGTCGA GGTGATAAGC AAGACTCTGA TAGTGAGGAA GAGACGGAGT 27101 TATAATAAAG AATGATCTCT TCTGAGTAGA AACTATCGAA TCAGAATTGC 27151 CAGCTTTTAC TCTCACCCAG TCCTATTTTT ATAGTTCAAA TTGCGGTGGT 27201 TTGTTTATTG TTTTTTTGAA TTTTTGTTAT ATGGTGGCGT CTAATATTGC 27251 CGATTAAGGA GGGCGCCCTT TATGAATAGA AGAAAAGCAA GATGGGTAGT 27301 GGCATTGTTC GCAATGACGG CGCTCATTTC TGTTGGGTGT TGTCCTTGGT 27351 CACAAGCGAA ATCAAGATGT TCTATTGATA AGTATATTCC TGTAGTCAAT 27401 CGTTTACTAG AAGTTTGTGG ACTTCCTGAA GCTGAGAATG TTGAGGATTT 27451 AATCGAGTCC TCGTCTGCTT GGGTACTGAC TCCTGAAGAA CGTTTTTCTG 27501 GAGAGTTAGT CTCTATCTGT CAGGTTAAAG ATGAGCATGC TTTCTATAAC 27551 GATTTGTCTT TATTACATAT GACTCAGGCT GTGCCTTCGT ATTCTGCAAC 27601 GTATGATTGT GCTGTAGTTT TTGGCGGGCC TTTGCCAGCG CTACGTCAGC 27651 GCTTAGATTT TTTGGTGCGA GAGTGGCAGC GTGGCGTGCG CTTTAAGAAA 27701 ATCGTTTTC TATGTGGAGA GCGAGGGCGC TATCAGTCTA TTGAAGAACA 27751 AGAGCATTC TTTGATTCTC GGTACAATCC TTTCCCTACT GAAGAGAACT 27801 GGGAATCTGG TAACCGAGTT ACTCCCTCTT CTGAAGAAGA GATTGCCAAA 27851 TTTGTTTGGA TGCAAATGCT TTTACCTAGA GCATGGCGAG ATAGTACTTC 27901 AGGAGTCAGA GTGACATTTC TTCTAGCAAA GCCAGAGGAA AATCGTGTGG 27951 TTGCGAATCG TAAGGACACC TTACTTTTAT TCCGTTCTTA TCAAGAAGCG 28001 TTTCCGGGAC GCGTGTTATT TGTAAGTAGT CAACCCTTTA TCGGTTTAGA 28051 TGCTTGCAGG GTCGGGCAGT TTTTCAAAGG GGAAAGCTAT GATCTTGCTG 28101 GACCTGGATT TGCTCAAGGA GTCTTGAAGT ATCATTGGGC TCCAAGGATT 28151 TGTCTACATA CTTTAGCGGA ATGGTTAAAG GAAACGAACG GCTGCTTAAA 28201 TATTTCAGAG GGTTGTTTTG GATGATTCAT GGATCTTAGA GGTTAAAGTC 28251 ACTCCAAAAG CCAAAGAGAA CAAAATTGTA GGCTTTGATG GACAAGCTTT

28301 GAAGGTCCGT GTTACCGAAC CCCCAGAAAA GGGTAAGGCC AATGATGCTG 28351 TAATTTCTTT ATTAGCAAAA GCTTTATCCT TACCGAAGCG TGATGTCACT 28401 TTAATTGCAG GAGAAACTTC TCGAAAGAAA AAGTTTCTTC TTCCTAACAG 28451 AGTTCAAGAC ATTATTTTTT CTTTGCATAT AGACGTATAG CCTAACTCAC 28501 AGCTACGTTT TTCCCTTTCT CAGATTTTTC CAATTTTTTG CTTTTAAAAG 28551 ATAAGAACTG GTTGCGTTCT GTTTTATGAA GCTTGATTCC TAAGTACTTG 28601 ATGATATCTT CATTAAAGGT GGTTGTAGGT GACAGGCGTT GAGCAATGAT 28651 TTTACGCAGG CTGTCCACAT CGTGATTGTT ATAGAGTAGG TGGTGCACAA 28701 TTTTTGCGAT TTGTTTTCCT GATTTTTTGT AATCCACGCT ACAGGCAATG 28751 CAGGCTCCTT CGGAAATTAA GGAGGTATCG TCGGTAATGA TAGGGATTTT 28801 CTCTTTGAGG ATTTCCTGAA GGAATGCGGT GCCTTCTTTA TGAGAAAGTG 28851 GGGAGAGGG AATGAAGATA GCTGAGGGGC GCTTGTCGAT AGCCTGGCGT 28901 ATCCGGGTTT TGAATGTACT GCTTGTAATA GAGATCTCAA TGACCTCAAT 28951 TCCTGAAGCA TGGAGTTTCT TAACAATTTC TTTTTGGAGA TCTGAGGGGA 29001 AAGGTTCGGA GGGTTTTAAA TACACGATAG ATTGTGCATT GGTAGCTACG 29051 GCTTGTATAG CAAAGCAGTA TTGATTGATG TCTAGAGTGT CATTCACTCC 29101 GTAGATATTC ATTGTGTTTT TAGGAAGGGT TAGGCTTTCG CGATCAGGAA 29151 CAGCGGCATA GATCACAGGT TTCTGTGTTT CAATGTGGCT CATGACCTTC 29201 GTAGCAATAG TTCCTAAGGT GACAATCGCC ACGACATTTT TATCGGTATG 29251 TAAGGAGCGA GCAATTTTCC TAGCCTTTAC GATACTGTCT TCAGCATTTA 29301 GGACAACAAT TTCAGGAAGG TTCTCAAAAT CTTTCAAGGT TTCTATACAG 29351 CTTTTACTGC AATCTTCTAA TAGGGGATGG GGAAAGGATA AGAAAATTGC 29401 GATTTTAGGA GAGGAGACGC TATCTGGTTG AGAACCACAA GTGGCTACAT 29451 AGATGAAAGA GCAAAACAGA GAAAAGAAAA ATAAGTACTG AGAGAGTTTA 29501 CGTAGAATTG TCATGCAAGA ACCATCGTTT CTTTTAGTGA AGCGGTACAG 29551 AGGCGGTCAC AGGCTAAAGC GATATTTTGT GGTTGTCA GAGCGGAGAA 29601 ACGAACAAAT CCTTGTCCAC AGGAACCAAA ACCGTGGCCG GGAGTCACTG

29651 CAATATGATA CTGATGTAAG AAGAAATCAA ACGCTTCTTC ATCAGAGATT 29701 CCTTCAGGGA GTTCTACCCA AAGGTAAGGG GCATGATCGC CACCATGAAC 29751 TGAGAATCCT GCAGTTTCTA AGCTTTTTTT AAGTTTCTGA GCATTGGTTA 29801 GATATAAAGA GATGGCGGGA GGTGTCGGAA ATAAATCTAG GCCGTAATAC 29851 CCTGCTTCTT GCATGAGGAG AGATGCTCCG TTAAATGTAG TCGCAAAGAG 29901 CCGTTTCCAA TCGTTGATCA TAGGTTCGTT ATTGTCATAG GTGAGTTCTT 29951 TAGGGATCAC GTTCCAGGCA AGGCGCATGC CAGTAAAGCC TAATGATTTA 30001 GAGAAAGAGT TGATTTCTAT AGCACAATAT TTTGCTTCAG GGATTTCGAA 30051 GATGCTTTTA GGTAGGCTAG GATCTGAGAC AAAGGCGCTA TAGGCCGCAT 30101 CAAAAATAAG AACGGTTCCG TGCTGATTCG CGTAGTTCAC AAGTGCTTGG 30151 AGTTGTTGAA AGGTTAGAAC TGTTCCTGTG GGGTTGTTAG GATAGCATAG 30201 ACAAAGAATG TCTAGGGATT GTTGGTTCGG AAGTTCTGGA ATAAACCCAG 30251 TTTCTTTTCT GCATGCTAGG GGGATAATGT CGCGGATTCC TGTAATGTGG 30301 GCAATGTCTC TATAAGCTGG ATAGACAGGA TCCTGTAGAC CTAGAGTCTT 30351 TTCTGAGCCA AAAAAAGAAA AGAGACGGAA GATATCAGGT TTGGCACCAT 30401 CCGAAATAAA AATCTCTTCA GGGGAGATTC TATTTTCATA GACTTCAGAG 30451 GCAATTTTTG TGCGTAATTT TTCTAATCCG GTTTCTGGGC CGTACCCACG 30501 ATAGGTCTCT TGTTTCTCTT GAGAAACGCA GAACTCTTTG ATTGCCTGAG 30551 TAATAGAGCG GCAGAGAGGT TGTGTCGTAT CTCCGATAGA AAGATCTATG 30601 ACAGAGATTT CTGGATTCTC CTTGCGAAAC TGAGCAAGCT TTTTACTAAT 30651 TTCAGAAAAT AGATACTGAG GCTTGAGAAG AGAAAAGTGG GGATTTCTAC 30701 GCATAGCGTG CTCCAGGTAA GGGCGTGATT CACTTGGAAG ACATCTTAAG 30751 AAAGCCCCAG CTTTTTGGAT AGCCATTTTC TGATTTTTCT TTAACCGCCT 30801 CTATAAAGGC AAGGAGTGCT CTGTAGTTTC GGTTAATTAT TAAGATATAT 30851 TTATAATTCT GTTTATTAAA AGTTTTTTAA ATCTTTTCTA ATTTGCTCAC 30901 TATAATTAAA GGATAAGATT TGAAAAAATT TTTTAGGTAA TTATGATAAG 30951 GGTCAATCCT TATGGAAGTT ATAGGGGTAG GAATCCTTCT CCAGAAGATG

31001 GGAAAAAGGA TGTACCCCTT TCAGGGAACT CTCGCTTGCA TCGTCGTGGT 31051 GGGATTCGTA GAAAGCATAA GAGTGCTTCA GTTGGGGTGA CCTCGGGTTC 31101 TAAGACGGGG AAAGCTTCTT TAGAGAAGAA GGTCAAAGGC ATTTCAGAAG 31151 CCCATTTCAA ATAATCCAAG ACAGAAGGTT CTCATTCGAA GACAAGCAAA 31201 GGATTCGTTG GCAGATTTGT TCAATGGATT AGAACGTTTA CAGGACGTGG 31251 AAGCAAGAAG CGTTCTCCCT CAAGTTTTTC TCCAACGCAC CCTTACATAC 31301 GTTTGCGAAC TTACACACGC AGTCCAAAAC AGAGTGGTGT AGAGAGAAAA 31351 CAAGAAGATG CTGAGACCTC ATTTATAGAG ACACCCAAAG GGATCTTGAA 31401 AAAGCCTGGA AACAAAGACC CCAAAGGCAA GCACGTCCAT TGGAAAGACA 31451 GCTAATCCGG ATCAGAGCTG AGATCCCTCC TTCTCTTCAT TAAGAGGGTT 31501 CAGAGTTTTA GGTCTCTGCG GTAGAAGGGA GGGAATCCAT AAGGAGAAGA 31551 CCAGTGGATT TTCAACGGAG ATTCTAGAAG GAAACAAGCT GATTCTACAT 31601 GGAGATGTTT GTAGGTCTGT TGTGAATTCC TCTCTTCACA GATCGGAAAG 31651 ATCAATGAAA AGAGTTAGGG TCTCTTAGTT TTAAGAATTG CTATACAATA 31701 TTGCGAAATA AATACTTTCT TCTCTATACA CTGTCTTTAC GATGAAGACA 31751 GCTTTTCACT CTTGCTATTC TTGGTTTTGT TGGCTCTTTA GCTTCTTGGT 31801 ACTCTTTGTG GGTGGCATCG CTGGGGGAGA GCCTTTGTGC CCCGATTGCA 31851 AATACGAAAC TAAGTCTGTT TTACGTTCGG ATCAGCTGCC GGATCATCTC 31901 TGGAACTATG AAAACGACTG TTATCTTACA GGTTATGTGC AGTCTCTTTT 31951 GGACATGCAT TTTTTAGATA GCCGTACGCA AGTTGTTATT GAGAAGAATA 32001 GAGCGTATCT TTTCTCTTTG CCTGTAGATT CGAGTTTATC AGAAGCCATT 32051 ACCAACTTTG TTAGGGATCT TCCCTTCATA TGTGCTGTGG AGATTTGCGA 32101 GCGTCCTTAT GGTGAATGCA TAACGAGATC TTCTGCGGAG CGTCCCTTAC 32151 TCCCTAAAGA GAAAACTTTA GGAATGCCAA TTTTCTGCGG CAAAGAAGGG 32201 GTATGGTTAC CTCAAAATAC CATTTTGTTT TCTCCTTTGA TTGCAGATCC 32251 TCGTCAGGTT ACCAACAGTG CTGGCATTCG TTTTAATGAG AAGGTCGTGG 32301 GGAATCGGGT AGGTGCTACC ATCTTTGGGG GAGATTTTAT TCTCCTGCGT

32351 CTTTTTGATG TTTCTCGATT CCATGTAGAT TGTGATTTCG GAATTCAAGG 32401 AGGAGTCTTC TCAGTTTTTG ATTTAGATCA TCCTGAATCG TGCATGGTAA 32451 ATTCAGATTT CTTTGTTGCC GGACTCTGGT CAGGGGCTAT AGATAAATGG 32501 AGTTTTAGGT TTCGATTGTG GCACCTCTCG TCCCATTTAG GAGATGAGTT 32551 TATTCTTACG CATCCAAATT TCCCAAGATT TAATTTGAGT GATGAGGGCG 32601 TCGATCTCTT CATTTCGTTT CGTTACACAC CACAGATCCG CTTGTATGGC 32651 GGCTGCGGTT ATATTGTAAG TAGGGATCTT ACTTTTCCTG AGCGGCCGTT 32701 TTACTGTGAA TGGGGTGCGG AACTCAGACC TTTTGGTCTG AGAGAAGGAA 32751 ATCTCCACGC ACAACCGATT TTCGCGATGC ATTTCCGTTG TTGGGAAGAA 32801 CAGAAATTTG GCTTGGATCA AAGCTATATT TTAGGCATGG AGTGGGCCAA 32851 ATTTCAAGAA ATCGGAAGGA AAATCCGTGC TGTTTTAGAA TATCATCAGG 32901 GATTTTCTAA AGAAGGCCAA TTCATTCGTG AACCGTGTAA TTACTACGGT 32951 TTCCGTCTTA CCTATGGATT CTAAACGATA ACAAAACCAT CTTCAGGGAC 33001 AGGGTGTTCA GGTCCTATGG GCAGCGTATG ATTGAGATAT CTGCGGAAGA 33051 TTGCAATGCC TGCCTTTGCT GAGGATAGGC AGAATAGGCA GTTGTGTACC 33101 CATTCAGAAC CTCGAATGGT TCCTACAGCA TGATTGGAAT TATACAAAGC 33151 TGTGATCTTG GATTTCCAAT AGTCTACAGG TCCGATTAGG AAGACGGGAA 33201 CAAGAGCTTT TTTCCCTGTT TTGAGACTAA TAAGCTCCAG AAGGAGTTCG 33251 AAATCGGTTC CCATGCCTCC GATAACAAAT ACAGCAAGGT CGACATGGAA 33301 GTCGGCCTGA CGTTCTAAAA GATCAGGAAT AGCATAGCTC ATTTTAGCTT 33351 CTACATAGGC ATTCGTGGTA TCCAAGCTAA TTAGATTCCC ACAAGAGAGT 33401 ATGGAGAGTT CTGTAGCTAC ACGATTCGCG AGTTCCATAG CTCCAGAACC 33451 CCCTCCTGTA AGGATTGCTA ACGGTGTCTG TGGTGGAAAT TCTGGGATCG 33501 TGAATTGCTG AGAAAGAGTA TGCATTCCTG TCAGGAGCTC ACGGAGAAAC 33551 TCATCATAAT CCCCAGCGAT TAGGCAGGAA CCATGAATTC CTATAAAGTA 33601 GGATTGAGCA AACTGTTCAG CTTGATGTTT AGGGACAAAC ATGCCCACAT 33651 CTTTATTTCT GCGTTTGATG TATTGTAAGA GTCGTTTCGA TTCTAAGTCT

33701 GCCCAAAATA CAGAAATTCC TGCAAAATAT AGATCGAGAA GGAAAGAGCG 33751 ATCTCGATTC GAGAAAAACT CTCCAGAAGT GGGAGAGGGA ATCTGAAAAT 33801 AGATATGTTG CAGGTAATAG CGAGAGTAGT TAGAGAGGAA CATGCCCTTC 33851 AGCGATGCTG AAGGGAAGTA GCGGGAAAAT AAAACTCCTT GGCTTGTGAT 33901 ATGATCTGTT TCCATGGCTT TTAAAAAAGG GAAACAAGGT TGGTCTTCAA 33951 TGTGCTTTTG AATTTCCCTA GCATGTCTTT CATCTGATGG GGAGATTCGA 34001 GGTTTGATGA TCCAAGAGTC TTGGGAGAGC TCAAGCAGCT CACTACCTTT 34051 GGAGATAAAC ATCGCAGCTT GATCTTCGCC TTCCGGTATG GATTCAAAAA 34101 CACGAAATAC CTCTTGAGGA GATTCTAAGG TTTCCTGGAG CATATCTCTA 34151 TAGAAGAAAA ACGAATGCTC TTTGTAAGGC TCAAGAGTAA AAAATTCTAA 34201 AGGTATTCTC TCAATAGGTT CTGAAGTGCT GCCGTAGAAT TCATAAATAT 34251 CTCCAGATTC TTGTGTGGTA GGTTCGAGAA TATCCGCTGC GGTGTGACGA 34301 AGCCCTTGGG GGAGTAAGTC CTGAACGACT CTTGCAAATA CGGTTCGGAT 34351 GTGCAGAGGC TCTGTCTTTA TGAGAAGAAT TTTATGATCT TCGGGAACGG 34401 GAGGACGATC TGTTACCATT TGATACAAAG GAAGAAACTT ACGTATTTTT 34451 AAATGGGGAC GCGTGAGTGA TTTGCTCATT AAGGGAAGGA ACCCATAAAT 34501 TGTCTCTTCG TAACAGATTG TTCCTGGAAG GATCGGAAGG AAGACAACAA 34551 GCCGATCATT AATGATCTCT AGAGTGATGA AGTGCTCAAG TTTTTTCCCA 34601 AAGCGTAGGA GCGGAGATCC TGTACGGTCT GTGTGCGTAA ACATCCTGTT 34651 GAGATAACAA GGCGAACGTA CGAGTCGGCG ATCATCAGCA GCAAAGAGCT 34701 TGCAGACAAA ACTTCCAGGC TCTAGGAGCT CCAACATAGC AGTGGCTATA 34751 GGATCTTGGC TCATGAAGAG AACGTGTAGA CGAGCTTCTT TTCGGGCTTT 34801 ATTTAGCTCC AAGTGGTTTA AAACGGCTTC GACACCTAGT TGGGCTAAGG 34851 AACTTTTAA ATTTACTTGT ATACACTGTT GAGGCAGATG AAATCCAAGA 34901 AAGTACGCAG GAATATTCTC AATGAGGACC TCTCCTTCGT AAATGTGGGG 34951 CGAGAGTTTT TTCAAATGGG AAACGAGTCG TCCGTCTGGG GAGGCTGCAT 35001 CATGATGCGC GTGGAGTAGG TTATACATAA TTCATCGTAT TTTAAAGTTA

35051 AAAGCAATAG GTAAGGTCCC TCTTTGCGCT TCTAAGACTA GCGTTTCCAG 35101 AAGAATTTCC TACAAGGACT TCATTCTGCT TGTTGTTATA ATTTTGTGCA 35151 ATTTGTCAAG GAGGGGGTCT TGGAACGAGA AAATTGTGTT GTGGAAGGGA 35201 ATTTTAGGAA AAAAGTTTCT TAGAAAATAG GGCAGAAACT TTCTCTTGGA 35251 GAGATTGAGT AAAATGTAAA GAATAGTCTT TGCAATTGAG AATTATTTCT 35301 CTCTGGTTAC AATGGAGGAT TGGCTAAGGA GGATAGTAGG TATGCAGATT 35351 CCAAGAAGCA TTGGTACTCA CGATGGTTCT TTCCATGCGG ATGAGGTCAC 35401 AGCGTGTGCT CTCCTTATTA TTTTCGATCT TGTGGATGAA AATAAAATTA 35451 TACGCTCTCG AGATCCTGTC GTATTATCGA AATGTGAATA TGTTTGTGAT 35501 GTCGGTGGTG TTTATTCTAT AGAAAACAAG CGTTTTGATC ATCATCAAGT 35551 CTCTTATGAT GGATCTTGGA GTAGTGCAGG TATGATTCTG CATTATCTTA 35601 AAGAGTTTGG TTATATGGAT TGTGAAGAAT ATCATTTCCT TAACAACACT 35651 TTGGTACATG GTGTGGATGA ACAAGATAAT GGCAGATTCT TCTCTAAGGA 35701 GGGATTTTGT TCGTTTTCTG ATATTATTAA AATTTATAAT CCTCGCGAGG 35751 AAGAAGAAAC TAATTCGGAT GCGGATTTTT CTTGTGCTTT GCATTTTACC 35801 ATCGACTTTT TGTGTCGGCT AAGGAAGAAG TTTCAGTATG ATCGAGTTTG 35851 TAGGGGGATT GTCAGAGAAG CCATGGAAAC CGAGGATATG TGTTTATATT 35901 TTGATCGTCC TTTAGCATGG CAAGAAAATT TCTTTTTTT AGGGGGAGAG 35951 AAGCACCCTG CAGCTTTTGT TTGTTTTCCT TCCTGCGATC AATGGATTTT 36001 ACGAGGGATT CCTCCGAATT TAGATCGCCG TATGGACGTT CGTGTTCCTT 36051 TCCCTGAGAA TTGGGCAGGT TTGTTAGGTA AAGAGTTGTC CAAAGTATCA 36101 GGGATTCCTG GGGCTGTGTT CTGCCATAAA GGTCTTTTCC TTTCTGTATG 36151 GACAAATAGA GAAAGTTGCC AACGTGCTTT GCGGTTAACG TTACAAGATC 36201 GAGGGATCAT ATGACAGTAT TCAAACAAAT TATCGATGGA TTGATAGATT 36251 GTGAAAAGGT ATTTGAAAAC GAAAATTTCA TAGCTATAAA AGATCGTTTT 36301 CCTCAAGCTC CTGTTCATCT TCTTATCATT CCTAAAAAAC CTATACCACG 36351 ATTTCAGGAT ATCCCAGGGG ATGAGATGAT TTTAATGGCA GAGGCTGGAA

36401 AGATCGTGCA AGAGCTTGCT GCAGAATTTG GAATTGCCGA TGGGTATCGT 36451 GTGGTTATCA ACAACGGTGC TGAAGGAGGA CAGGCGGTAT TTCACTTACA 36501 TATTCATCTT TTAGGTGGGC GTCCTTTAGG TGCTATAGCC TAATTTTCTT 36551 TTGTTTCTGT GGATCCTTGT TCGGCTCGGA GTCCCTCCGT TATCAATTGT 36601 TGATCCAAGA TTTTGCAAAA GTTTCAGAAG AGGGCATAGG CCTTTTGGAG 36651 TCTAAAGAGT ATTCTTTACT TCAGGCTAAG CTAGTTTTAA GGGCTCTGGC 36701 TCAAAATTCT TCTTTTGATG ATTGGTTTAG AAGTTTTAAG AAGTGTCAGA 36751 TTTCCTATCC AGAGTTAGCT CATGATCGCG ATGTCTTAGA AGAATTTGGG 36801 ATTCAAGTTC TGCGTGAGGG AATCGAAAAT CCTTCCGTGA CCGTTCGTGC 36851 TGTGAGTGTC CTTGCTATTG GGCTTGCTAG AGATTTTCGC TTGGTCCCTC 36901 TCCTGCTCCA AAGTTGTAAT GATGACAGTG CTATTGTTCG ATCTTTGGCT 36951 CTTCAGGTTG CTGTGAACTA TGGCTCTGAA AGTTTAAAAA AGGCCATTGT 37001 AGAGCTTGCC CGTAATGATG ATTCTATTCA TGTTCGGATT ACAGCATATC 37051 AGGTGGTCGC TCTTTTACAG ATAGAGGAGC TATTGCCATT TTTAAGAGAG 37101 CGTGCTGAGA ACAAACTTGT AGATAGTGTA GAACGTCGAG AGGCGTGGAA 37151 GGCTTGCTTG GAACTCTCTT CTCAATTTCT AGAGACGGGT GTAGCTAAGG 37201 ACGATATTGA TCAAGCGTTG TTCACTTGTG AAGTGTTGCG TAACGGTATG 37251 TTGCCAGAGA CTACTGAGAT TTTTACAGAA CTCTTATCTG TAGAGCATCC 37301 TGAAGTGCAG GAGTCTCTCT TACTTTCTGC TTTAGCTTGG AGTCATCAGC 37351 TACAGAATCA CAAAGAGTTT CTTAGTAAAG TGCGCCATGT GATGTGCACT 37401 TCTCCATTTG CAAAAGTACG TTTTCAAGCT GCTGCACTTC TCCATCTGCA 37451 TGGAGACCCT TTGGGCAGAG ACTCTCTGGT TGAGGGCTTG CGCTCTCCTC 37501 AACCTCTTGT GTGTGAGGCA GCTTCGGCGG CTCTCTGCTC TTTAGGAATC 37551 CATGGAGTCC CTTTGGCAAA GGAGCATTTG GAGAGCCTTT CTTCTCGAAA 37601 GGCTGCTGCG AACCTCTCCA TTTTGCTTCT TGTGAGCCGT GAAGATATTG 37651 AAAGAGCTGG AGATGTGATT GCTCGCTACC TCTCCAATCC TGAAATGTGC 37701 TGGGCTATAG AGTATTTCTT ATGGGATGCA CAATGGAATT TACGTGGTGA

37751 TACCTTCCCT CTATATTCGG ATATGATTAA ACGTGAGATT GGTAGGAAGC 37851 GCAACGTTCC TTTCAGGACA GCAAGCTCAG GGATGGAGCT TTTTTTCTGG 37901 AATGTTCTGG GAAGAGGGAG ATGTGAAAAC TTCTGAGGAT TTGGTTACAG 37951 ATGCTTGCTT TGCAGCAAAG TTGGAAGGAG CGTTAGCCTC GCTATGTCAG 38001 AAAAAAGATC AAGCTTCCCT ACAGAGGGTC TCTCAACTTT ATAATGACAG 38051 CCGTTGGCAA GATAAATTAG CAATCTTAGA GAGCGTTGCT TTTTCTGAGA 38101 ATCTTGATGC TGTGCCTTTT CTTCTAGACT GCTGCCATCA CGAAGCTCCT 38151 TCGCTGCGAA GTGCAGCAGC GGGTGCTCTT TTCTCTATTT TCAAATAAAT 38201 ATTAATAAAA TTATTCAAGA TATAGAAGAA AAACCACGCA GTTGTAATTT 38251 CTATTCTTA AAAATAATTT TTCAGACTGA CTTTATTCTT TCATTTTTAA 38301 GTCTTTGGAG ATAGAAAACT TTGTTATAGA TTTTTATCTG GTAGCTTTTA 38351 TAATTTATGA AGAGCGTAAG CTCAGAGCCT GTATGTCATG CACAACCTGT 38401 ATATAAAATC AGATTTGTTT TTTGAATCTC TATTCTCGTT AAGATTTCGT 38451 TATTCTGGAC GTTATCTCCA TCACCACTCC TAATTTTCCT AGCATTTCTA 38501 TCTTTAAGCT CAGCACCGTA GCTTGCTTAA AGGAAATATT TTTCATTTAG 38551 GTTGTGGAGT TCTTTATTTT ATGAATTTTT CATTATTTTT ATTTTTCCTG 38601 ATAGCTATTC AGGGAATCTG CTTGTACGTG GGACGTCGTG GTAGCAAAAA 38651 GGTAGAAGAT CGCGAGAGCT ATTTTCTTGC AGGAAGGAGT TTAAAAATCT 38701 TTCCTTTGAT GATGACATTC ATTGCCACCC AAATCGGTGG CGGTGTACTT 38751 CTTGGGGCTG CTGAAGAGGC CTTCTGTTAT GGTTATGGGG GGATTCTTTA 38801 TCCTTTAGGA GTCGCTTTAG GGTTGATTTT CTTAGGAATG GGGCCCGGGA 38851 AGCGGTTGGC AGAGGGATCG TTAACGACCG TAGTCTCTAT CTTTGAAGTG 38901 TTTTATGGTT CTAAAAAGCT CCGTAAGATC GCATTTTTAT TATCCGCAGG 38951 TTCCTTATTT TTCATCCTGG TCGCTCAGGT GATTGCTTTA GATCGGTTGT 39001 TTAGCAGCTT CCCTTTTGGC AAGTACGTAA CCGTAGCATT TTGGATTGTC 39051 TTAGCATCCT ATACCTCAAC AGGAGGGTTT CGCGGGGTCG TACGTACTGA

39101 TGTGATCCAA GCAGGATTTC TTCTTATTGC GGTGCTCGTC TGTGGTGTTT 39151 CTGTATGGCT CTCTGTCCCT AAATCCTTGT CTGTGTTGGA TCCTTTCCAA 39201 TCACTTCCTT GTGCGAAGCT TTCCAATTGG ATATTCATGC CTATGCTCTT 39251 TATGCTTGTT GAGCAGGATA TGGTGCAAAG GTGTGTGGCT GCCTCCTCTC 39301 CAAAACGCTT GCAATGGGCG GCTGTAGGCG CAGGCCTTGT TCTTCTTCTT 39351 TTTAACTTTA TCCCTTTATT TTTAGGTTCT TTAGGAGCTA AAGCAGGCCT 39401 TAAAGCAGGA TGCCCTCTGA TTGATACCAT TGCATATTTT TGCAATCCCT 39451 CACTAGCAGC TGTGATGGCT GCTGCCATCG GCGTTGCGAT TCTCTCTACC 39501 GCGGACTCTC TTATGAATGC TGTAAGCCAG CTAATCGCTG AAGAATACCC 39551 TACGTTGAAA GCCCCTTATT ATCGTTATTT AGTATTGGGT TTGGCGGTTG 39601 CAGCTCCTCT TGTTGCTATT GGTTTTACAA ACATCGTAGA TGTCTTGATT 39651 TTAAGCTATA GCCTGTCAGT GTGTTGTCTT TCAGTCCCTG TGGGTTTCTA 39701 TCTTCTAGCT CCTAAAGGTC GCCGTGTGAG CGGAGCTGCT GCTTGGGCAG 39751 GAGTGCTCGT TGGTGCTCTG GGCTATGGAT GGGTTCAGAT AGTCTCTTTG 39801 GGGATGTTTG GGGAGCTATT GGCTTGGGTA GGTTCTCTAG TCGCCTTTTC 39851 CTTTGTAGGA TTTATTGAGA TCACTTGGAA AAACAAAGTC AAAACGCAAA 39901 CTTAGATAAC CACTGCATGA GAAGATATAA CTAAAATAGA TCCTGAGTTG 39951 TTTAGGTTTC TCTTAGATCT GATAGGTTGC GCTTAGTAAG AGATCGTCAG 40001 TTTTTTAAGT TGTGTTTAGA ATCTGATACC TCTCCTTCTT TTCCAAGAAG 40051 AAGAGGGTT CGTTTTATTT TTTATTATTC ATTCGTAGGG GCGGGAAGCT 40101 GTTTTAACTG ATAGAGCAGG TCGATAAGAG AGGAGAGCAG GGCTTGATAG 40151 AGCGTCTCTT CAGAATGGTG AAGAGTGCCC TTAAGAAGCT TTCTGCAATA 40201 CTTTAATAGC CAATACAGAG CGCAAAGTAA ACACTGTAAA AGGTAAAATT 40251 TACACGCTTG TTTTATCATA AACCTCCAAA GAAAGACGCG TGTTGGCGTT 40301 CTTCCAATTG GCTGTTAGAT AGTGAAAATA GTATTTACTA TTCAATAAAA 40351 ATATTTATAT TAAATATAAT AAAACAAAT TTCTAATAAA CTTTTTAAAA 40401 GTATTTGGCT AAGTTTAATT TAAGAGTTCT CAAATAAAAG ATTTTTTAGT

40451 CTCTTTTATT TGAGAATCAT CAACCGATTT CAAGAAATCG CATGAGCCCG 40501 TGATTCTGAT GGGCAGAAGT GCTTTTTAGC AAATAGGTGA AGCTCCTCGG 40551 GCGTGATCTG TTGCTTCGCG TCACAAAGAG CCTTTTCAGG GTGTGCATGC 40601 ACTTCGATCA TCAGACCGTC GGCACCTACC GAGAGACCAG CAGAGGCGAG 40651 AGGAAGAACT AGAGAACGCT TCCCCGCTGC GTGGGAAGGA TCTACAATTA 40701 CAGGGAGAGA AGAGATCTCT TTAAGGAGAG CCACGGTATT GAGATCTAGC 40751 GTGTAGCGCG TAGAGTGCTC AAAGGTACGA ATTCCTCGTT CACAAAGGAT 40801 TACCCCAGGA CAGGAGGGAG AAGAAGCAAG GATGTACTCC GCTGCGCATA 40851 GCCACTCTTC AAGAGTAGCT GCTGGACTGC GTTTTAGGAT AATCGGACGA 40901 TGTGATTTGC TGACCTCTTG TAAAAGAGGG GTGTTATGCA TGTTTTTGGC 40951 TCCGATACGG AGGATATCCA CATGTTCGGC AGTAATTTCA ACATCTCGGA 41001 CATCTAAAAC TTCGGTTTCT GTAGGGAGAC CATGGATGCT CTGTGCTTCC 41051 TTATGCCAAA GCACACTC TTTCTCCCAT CCTTGAAACG AAAATGGGCT 41101 TGTCCGTGGT TTTCTGATTG ATCCTCGGAA TACCTGAGCT CCTGCTTCTT 41151 TAACTGTAAG AGCTGAAGAG ACTGTATGCT CGTAACTTTC TAAGGTGCAG 41201 GGGCCTGCGA TCAGTATTGG CGATCCTTCT CCAAACGATA GATTTGGAGA 41251 AATAGGAACG GTATGGACCT CGTCAGGATG CTGTTTGAGG GTGCGCGGTA 41301 GGGGATAGGT AAACGTAAGA ATAAGTACCT CATGCAAAAC TAGGGATTTG 41351 AGCTGTCCGG TTTTGAGTTC TGATCTTCAG TTCTAGGGTT TCTAGGCAAG 41401 AGACAACCGT AGTGGTTGGG GAAGCGGATT AAGAAGATAC GAGCGTCTCT 41451 TTCCGGAGAA ATTGTATTTA AAAATCCATG CATCGCAACA AAGTTTTCAA 41501 TATTTTCTTC TGAGACAGTA TCTCGATCAG ACATGATCAG ACAAACAGCA 41551 AACGGCTGCA AAGGTACAGT AGTCATAGGA GCGACGTCGT TAGCCTCAGC 41601 TTGCTCTTTG ACGTCATCCA AGATATCACC GAGTTCATTA GAACTTAGGG 41651 AGTGGAGGTA GCCTATCAGT GATTGTCTAA AACCAGAGCG ACAGAAACCA 41701 TCAGATGCTG TGACCGTGTT TTTTAAGAGA TTACAAATCA ACTGATCTCG 41751 AGCTCTAGAG TATTGATTTA TAGTCGCTGT AGCTAGAGGG AAACCGTATT

41801 TTAACGAAAG TAAAAGCTTA TCTTTTACTT GTGTAAATCT ATGAGAAGCC 41851 TCTCTTTGGA AAGCGTCTAA TCGAGCAGCT AGCGGAGTCT TTTTTTGTAG 41901 AAGTTTAGGG TAGTTGTTAA GGAAACAAAA GATTAAAAAA GATTTCGAAA 41951 ACTGATCACT TAAAGCTGTG GGAGAAGAAG GCGAACCCAG TTTTGATCCG 42001 AAGAGATTTA ACCGTGTTAA AAAAACGTTC CACCCATTCC TAAAGGTACT 42051 TTCGATTTCT TTTTCTAAGA GTATGGCCTC GGTTTCCAGT TGATAATTTG 42101 GGAAAGTAGC TCGTAGAAAC TGCATACAGT TCTTTTCAAG TATGGGATCT 42151 AAAACCATCA CTTCAGGATG TCGTGACCAG AACTTGCAGA ACGTCTCTTT 42201 ATCTTTATTA GAAGGGGTTG AAGGATAGCG TGCTGGAGGC TTGAGTACTA 42251 GAGACCCTAG AATGGAACGT AGAAGTTTTT GGAGTTTCTT TTCGTTTTCT 42301 ACTTTTTAA AGGCATATTG TAGAAAAGGA AGTAGAAGCA TTTTTAACTC 42351 TTTAACGGCA TCTTTAGTTT CTGGACGATT ATTCACTTGA GGATGCTGTA 42401 AAATAAAGAA TAAGAGCTGT ACTGCCTGTG TATAGAGGAG GTCCTGCTCT 42451 TGTGAGAGAC CACTGCCTTC ATCACGCGTT AAACGTAAAC TTCTTGCCGA 42501 GCAGACAGTA TTGAAGTGCT CTTTTAATTC AAGCTTACTA TGAATTGCGA 42551 CTAATAGCTT GTTTACAAGA TTTGAAAGTA AATCCGGCTG ATTGCTTAAA 42601 GGAGGGGTTT CAAAGGATGT GGATATTACA TTCTTTAATA GGTTTGCTAT 42651 GCGTTTATGT AGTGTTTCGA GTTGAGGCAT AGAGCTCTCT AACGTCAAAC 42701 TCTCATGAGT CGTAAGGATA GTCATGATGT TAGAAGAAAT CACCTCTCGT 42751 TGCCAAGAAG AATCTAAATC TTCATTAAGG TAGCTGAGGG AGGCGTTACT 42801 TATTTGTAAG ATCTCTTGTA TAGAGCGTTG GGGGGGCGCG GACGCTATAA 42851 TGATATCCTC TAATGATTTG TAGAAAGTAG CGCAGAGTTT GTGAAATCTT 42901 ACAGAGGGC AGGTACATAG AAAAGCAAAG AAAGAAAGAG CTGGAGACCA 42951 AGGGAGATCC CCAGATCGTC TGTTTTCTAT AGTTGCAAGA AAGCGAGAGT 43001 ATTGACATCG GATTGCTTCA GGGAGGTGAT CTTTGACTAT AGCATTGAAT 43051 CTCTCAGTAT CCCACCCTGA ATCAGCAATG CGTCGACTCA GCTTAGCAAA 43101 AGCCTCTCGT ATTTCTCTTT CATATTGCTT TCGAAGTTGT TGATCTTCTG

43151 GAGACATCTT TTGAAGCTCT GCATCAGTAA GAAAGAGAGC GGGCAGATGT 43201 TCCAAGAGAA GAAAGAGTTT GTCTCTAGAT TGCATGTAAG GAAGTGCGGA 43251 ATGATCAATG AGATTTTGGA ACAATTCCGA ATAGGGACGA TTTGCAGCAA 43301 GAAATTCCAG AACACTGGGG AGGAGATCGT CCTGCATATC AGAAATAAAT 43351 AAGGCTCTCG CTGTTTCTTC ATTACTTGAA GCTGCGATTT GTTGGCGAAT 43401 CGCATATGCA GAGAGTTTTC TTAGAAAGGC TATCAGAGTT GCAGTGTGTT 43451 TCTGAGAAAG AATCACCCCG TCATAGAGGT CTATGAAGGA GCAATAAGTA 43501 CTGCATAACT GAAGGAGTTC AGCCATTTCT GCACAAAGAT TCGCATTTGC 43551 CGGAGAAGAG GAGGCGAAAG GCAGGTCAAG GAGACGTGTG GCTTCCTGCT 43601 CAAAGACAAT ATCATTTCTG CTGCTCTCTT CGTAGAGAGC AGATAGCCAT 43651 CCTACAGCAT AGGCACGATA AAAGCAGTTC CCATCTCCCG GTACATTCAC 43701 AAGGTAGTAA TTGTCATTTA GATAGAGAGC CTGTTCAAGA GAGAGTTGCG 43751 CAAGTCGCCG GTGTTGTTGA GGAAGATCCG GATTTTGTGC GATTTTCTTG 43801 AACTGTTTTA TTTGGAAATA CATCGGTTCG TTATCAATCC GATTGGGATA 43851 GGAAGCTACG AAATGAGGGT TAAAGTCGCC AAGTATTGGA TCAACTTGCA 43901 TGGCAGGAAG AGGCAGAGGC GCTCGTCTTA GTACCATGCT CACCATGCGA 43951 TTTAAGCATT GCCAATAGCC TTGACTAGAT GGGCGGAGAG GCATGGGGGT 44001 AGCCACCCGT ACGGGAGGAG GAGGGGCCTC TGGTGGTGGA GTCGGCTTTT 44051 TATCAGCAGG TTGTTTAGGG ACTTTTGGGC TCGCTGGTGA AGGAGCTTTG 44101 GGGGGAGGCG GGGGTGTGTC CTCTGGGGGC GGCGTGCCCG GCTTGGGAAC 44151 ATCGGGTTTT TTGTCTTCAC CATCCTTAGG CGGTTGTTTG GCAATTTCTA 44201 TAGTTTTTGG CTCTGGTCCT TTGGGAAGAG TGGGAGGCGT TGGCAAGCCT 44251 TCTTTTCTGA CAACCCGATG ATGCTTGTAG TAGTGAATCA GAAGAAGCAA 44301 ACCAAGAGTA ATGATATGGA GCAGAACGTA TCCTATGGTA CGTAGAATTC 44351 TAAGTAACAG AGGGTCTTTA GTATCAGTCG TTAAGTGGTA AAAATTATTC 44401 TTGTTATTGG GCGGGCAATG TGGTGGGGAA TTGGATACAT ACGTTCAAAA 44451 ATTGCTCGTT TTTTAATCAA AATTATTCAA AGTTAAAACT TTTTCGAGTT

44501 TGATGTATTT ATTTTTTAT GTAAACTTTA ACACATGATA GAATTTAGCG 44551 TATAGAGCGC AAACTTTCAT GATAAAACAA ATAGGCCGTT TTTTTAGAGC 44601 ATTTATTTT ATAATGCCTT TATCTTTAAC AAGTTGTGAG TCTAAAATCG 44651 ATCGAAATCG CATCTGGATT GTAGGTACGA ATGCTACATA TCCTCCTTTT 44701 GAGTATGTGG ATGCTCAGGG GGAAGTTGTA GGTTTCGATA TAGATTTGGC 44751 AAAGGCAATT AGTGAAAAAC TTGGCAAGCA ATTGGAAGTT AGAGAATTCG 44801 CTTTCGATGC TTTAATTTTA AATTTAAAAA AACATCGTAT CGATGCAATT 44851 TTAGCAGGAA TGTCCATTAC TCCTTCGCGT CAGAAGGAAA TCGCCCTGCT 44901 TCCCTATTAT GGCGATGAGG TTCAAGAGCT GATGGTGGTT TCTAAGCGGT 44951 CTTTAGAGAC CCCTGTGCTT CCCCTAACAC AGTATTCTTC TGTTGCTGTT 45001 CAGACAGGAA CGTTTCAGGA GCATTATCTT TTATCTCAGC CCGGAATTTG 45051 TGTCCGTTCT TTTGATAGCA CCTTGGAGGT GATTATGGAA GTTCGTTATG 45101 GGAAATCTCC GGTTGCCGTT CTAGAACCCT CGGTAGGACG TGTCGTTCTT 45151 AAAGACTTCC CTAATCTTGT TGCAACAAGA TTAGAGCTCC CTCCTGAATG 45201 TTGGGTGTTG GGCTGTGGTC TCGGCGTAGC TAAAGATCGT CCTGAAGAAA 45251 TACAAACGAT TCAACAAGCG ATTACAGATT TAAAGAGCGA AGGGGTGATT 45301 CAATCTTTAA CCAAGAAATG GCAACTTTCT GAAGTTGCTT ACGAATAGAG 45351 GGTATTCTTA TGGCAACCTC TGTTCCTGTA ACTTCATCTA CTTCTGTAGG 45401 AGAGGCTAAC TCCTCCAACG AAAGATTTAC TGAACGAACA TCGCGAATGT 45451 ATTACGCAGC TTTAGTCCTA GGGGCTTTGA GCTGTTTAAT TTTTATTGCT 45501 ATGATTGTCA TTTTCCCACA GGTCGGATTG TGGGCTGTGG TCCTCGGGTT 45551 TGCTCTTGGA TGTTTACTTT TAAGCTTAGC TATCGTTTTT GCTGTCTCCG 45601 GTCTCGTTTT AGGCAAGACT TTAGAACCTA GTCGAGAAGC GACTCCTCCA 45651 GAAATTGTTG CGCAAAAGGA GTGGACTACA CAACAAGATG TCTTAGGGAA 45701 TGAGTATTGG CGTTCCGAGT TGATTTCCTT GTTCTTACGA GGGGATCTCC 45751 ACGAATCTCT GATTGTTGAT TCTAAGGATC GATCTTTAGA TATTGATCAG 45801 AGTTTACAAA ATATATTGAA ACTTGAGCCC CTATCTACGA CACTTTCGCT

45851 GTTAAAGAAA GATTGTGTCC ACATCAATAT CATTTTACAT TTAGTGAGAC 45901 AGTGGAACTT ACTGGGAGTG GATCTTAGTC CTGAAGTCAC TGCGCACGCC 45951 GAGGAACTTC TACTCTTTTT GATAGAAGAG CAGTATTACT CTCCTGATAT 46001 TTTGAAATTG ATTCGCTACG GAGATGCTTT ACAAGCAACG TCTCCTTTGA 46051 TGGATTGGGC AGATTCAGGT TCCTTTAGTG TAGACGCAGA CGGGGTATTT 46101 AGCTGTCGCA GAGAAGAATG TTCTCCTGAG GATGCTTTGG CGCAATTCGA 46151 TCTTCTTTG GCGTTGGAAA ATCCCGACAG ACGCTTCTTA AAGGATTCTT 46201 TTCTTACCTA CATTTGGTCG TCTTCATTTT TTGAGAAGTT TTTACATCGC 46251 CATCTAGAGA GCTTGCAAAG AAAGCTCCCA GAGACAGCGA TCGATGTCGC 46301 CCGCTATGAA GCACAAATAC AAACATTTCT CTCTCGCTAT TTTCAGAAGC 46351 TCGATTTGAT AAACGCAATG TCCTTAGATT GGGGATATAA CTGTGCTGAG 46401 GGAGAAAAAT GTTATGAGAG CGCAAATCAA AGATTAGACA ACCTATTTAT 46451 TGCTTTTCT TCTTCTGTTC CTGCTATGAA GCGGCTCTTT GACAAATATG 46501 GTTCTGTGGT ACGGGTAGAT CGTAGGCAGA TTCGTGAGCA GATTCTTTCG 46551 AACACTGAAA TCTTAGAAAA TGAGTCAGGG TTCCTCTGCA GTTTGTATGA 46601 ATATCCTTTA TCCTATTTGA TAGATTGGGC TGTTTTGCTA GACTGTGTTC 46651 GCGGTACCGA AATCTCTCTA GAAGATCAGG CCGATTACAC CGTTTGTTTG 46701 CAAGGCTTGG ATTCTATGTT ATCTCAATTT GCGAGTCGTT TACAGTCTGG 46751 ACAAAAGTA TTGAATCCTA GAGATGTTTT AAGTGAACAG GCTGCGGTTA 46801 TGCTTGTTCA TGGCTTGGCA GCACAGGGCG TGTCGTTTCA AGGATTGAAA 46851 GCTTTGATGT ATTTGACAGC CGTTCCCCAA AGAATGTGGT TAGGAGCATT 46901 GCCTTTATTT GAATCTTTTC CTGTCTTTAA TCGGATGAAA GAATTTCTTG 46951 GGGAATCTCT GGGAGACTAG GTGAATTTGT ATCAAAGAAG GAACAAGATT 47001 GCATGTTAGG TTCTTTGCCA TGTTATCCTG GTGCTGGCAA TATTGAAGAA 47051 TACAAAAATA GGTATTTCTA TTGTCAGTTA TGTGCTGAGG TCGTTAGTCC 47101 CTATGTTGTT CCTGTTATTG TAGTTGATGT GCAAGGGGCT CCTCCTACAG 47151 GTATCTTGCA GGTCTTGCGT TGTAAGCAAC ATAAATTTCA AGGCCTACCC

47201 GTACATGGCC CCATTACTTC TTTATGGGCT TTGGAGCCCG TGGGTAAGGG 47251 AGCTCCGCAG CTGGAGTCTG CAATGTACGA GCTCTGTTCT CAAGTAAGGA 47301 ATTTGACAT CTGCTCTATT GTGAGTTGGG TCTTTGGTGG GTTGTGTATT 47351 TTTGCAGGTC TGATTGTCGG GGTAATGGTT GAAGCCCCTT TGATTGCGGG 47401 ATTAAGTGCT TGGGTGATTC CCTGTATCAT TGGAGGGGTT GGTGCCATTT 47451 TATGCTTGTT TGCGATCTTG ATGGCGTACT TGGGAAGAGG GAGAGTCCGT 47501 GAGTGGCTCA ATCTTTCACA CGAATATATA ACGCAATGTC ATTGTCGTCA 47551 GATACAGGCA CATTCTCAAA ACTATTCTGT GATCACAGAG TATCCTGCAA 47601 CCTGTGCATT ATCTCAACCG ATTACAAAGT TACCTAATGG ATCACGCAGA 47651 GATAACTAAG CGTGTTCGTC AGTTATTTCT CACATTTTCT CATGAATCTT 47701 TTACTGCGCT GCACGAGATC CCTCTCGAAA ATTTTTAAGG ATAGATACTT 47751 GGAAACTATG GTTTAAAAAG CTATAGAGGA TTCTAAATTG GGGTTCTAGC 47801 AACTTCTTGA CTTTAAGATC CAAAGTTAAG AGACTGACTA ATTATTTTTG 47851 TTTGCTTGTG TTTCCAGATG AGCAATTGGT ATGGTAAGAG ATATTCAGAG 47901 TGAATCTATA GGGAAATTAG TATTTTTAGG CACAGGAAAT CCCGAAGGAA 47951 TTCCCGTGCC GTTTTGCTCA TGTAGAGTGT GTCAAAACAC AGGGATTCAT 48001 CGTTTACGAT CTTCGGTACT CATTCAATAT CAAAACAAGA CTCTAGTGAT 48051 TGACGCAGGC CCTGATTTTC GTACGCAGAT GTTAGTTGCA GGGGTTTCCG 48101 AGCTCGATGG GGTATTTCTG ACCCATCCCC ACTACGATCA TATCGGTGGT 48151 ATTGATGATT TACGTGCGTG GTACATAGTC ACGCAGCGTT CGTTGCCTTT 48201 GGTCCTTTCT GCAAGCACCT ATAGATTTTT AAACAAGGCT AAAGAGTATC 48251 TCTTCGCCAC TCCGAATGTA GAGTCTTCAC TTCCCGCAGT TTTAGAGTTT 48301 ACAATCTTGA ATGAGGACTG TGGGCAGGAG GAATTTCAGG GCATTCCCTA 48351 TACTTATGTT TCCTATTATC AAAAGTCGTG CCATGTAACG GGTTTTCGTT 48401 TTGGAAATCT TGCTTATCTT ACAGATCTCT GTAGCTATGA TGCAAAAATT 48451 TTCAGTTACT TAGATAATGT AGAGACATTG ATCTTGTCTG CGGGTCCATC 48501 GGAAACTCCT ATTCCTTTTC AGGGACACAA ATCTTCGCAT CTTACTGTAG

48551 AAGAAGCCAA AGCTTTTGCG AATCATGCAG GGATAAAGAA TTTAATTATT 48601 ACACATATCA GCCACTGTTT AGAAGCAGAG CGTGACCAGC ATCCAGAGGT 48651 CACATTTGCT TATGATGGCA TGGAGGTCCT TTGGACACTA TAGATACGCC 48701 CGGGGAACAG GGTTCTCAAT CTTTCGGAAA TTCGTTAGGG GCCAGGTTCG 48751 ACTTGCCTCG TAAGGAACAG GATCCCTCTC AAGCTTTAGC TGTGGCTTCC 48801 TATCAAAATA AGACAGATTC TCAGGTCGTT GAAGAACATT TAGACGAGTT 48851 GATCTCACTT GCGGATTCCT GTGGTATTTC TGTTTTAGAG ACCCGTTCTT 48901 GGATTTTAAA AACACCCTCA GCTTCCACCT ATATCAATGT GGGGAAGTTG 48951 GAGGAGATCG AAGAAATCTT GAAAGAGTTT CCCTCTATAG GGACTTTGAT 49001 CATAGATGAG GAGATCACTC CATCCCAACA ACGGAATTTA GAGAAACGCC 49051 TTGGCCTTGT CGTTTTGGAT AGGACGGAGT TAATTTTGGA AATCTTTTCC 49101 AGCCGTGCCC TTACTGCAGA GGCAAATATC CAAGTCCAAC TTGCACAAGC 49151 ACGTTATCTC CTTCCTCGTC TTAAGAGACT TTGGGGGCAC CTATCTCGGC 49201 AAAAATCTGG GGGAGGTAGC GGAGGCTTTG TTAAGGGGGA AGGAGAAAA 49251 CAGATCGAGC TAGACCGTAG AATGGTCCGT GAGCGTATCC ATAAGCTGTC 49301 AGCACAGCTG AAAGCTGTGA TCAAACAGCG TGCGGAACGC CGTAAAGTAA 49351 AATCTCGACG AGGAATTCCT ACCTTTGCTT TGATAGGGTA TACAAATTCA 49401 GGGAAGAGCA CCCTATTAAA TTTGCTGACG GCTGCTGATA CGTATGTTGA 49451 AGACAAGCTA TTTGCAACTT TAGATCCCAA AACGCGCAAA TGCGTACTTC 49501 CAGGAGGCCG TCATGTCCTT CTTACTGATA CTGTAGGCTT CATTCGAAAA 49551 CTTCCTCATA CTTTGGTAGC AGCATTTAAA AGTACTTTAG AAGCAGCTTT 49601 CCATGAAGAT GTTCTTCTGC ATGTTGTCGA TGCTTCGCAT CCTTTAGCTT 49651 TAGAGCATGT ACAGACGACC TACGATCTCT TTCAAGAGTT GAAGATTGAA 49701 AAGCCTAGGA TCATTACTGT GTTGAATAAG GTAGATCGGC TTCCTCAAGG 49751 AAGTATCCCT ATGAAATTAC GTTTGCTCTC TCCTCTTCCT GTATTGATTT 49801 CAGCAAAAAC TGGGGAGGGG ATCCAGAATC TTCTTAGTCT TATGACGGAA 49851 ATCATTCAGG AGAAAAGTTT GCATGTGACT TTGAATTTTC CTTATACAGA

49901 ATATGGAAAA TTTACGGAAC TTTGCGATGC CGGGGTTGTG GCCTCGTCAA 49951 GGTATCAAGA AGATTTTTTA GTTGTTGAAG CGTATCTTCC TAAGGAGCTG 50001 CAAAAGAAAT TTCGTCCTTT TATTTCTTAT GTTTTCCCTG AAGATTGTGG 50051 AGATGACGAG GGTAGAGGGC CCGTCTTGGA GAGTTCTTTC GGGGATTAGG 50101 TAGTTTTCTT CTAGGACATC GAATCTTTGT TAGTGAGAAA AAGAGTGATA 50151 TTTTAAAATA GCCACTCATC GCTAAATCTA TTGAAGTCTC TAGAGGTATA 50201 TGACGGTTGC GGAAGTCAAA GGAACATTTA AGCTGGTCTG TTTAGGCTGT 50251 CGGGTGAATC AGTATGAGGT CCAAGCATAT CGCGACCAGT TGACTATCTT 50301 AGGTTACCAA GAGGTCCTGG ATTCTGAAAT CCCTGCAGAT TTATGCATAA 50351 TCAATACGTG TGCTGTCACA GCTTCTGCTG AGAGTTCGGG TCGTCATGCT 50401 GTGCGTCAGT TATGTCGTCA GAACCCTACA GCACATATTG TTGTCACAGG 50451 TTGTTTGGGG GAATCTGACA AAGAGTTTTT TGCTTCTTTG GATCGGCAAT 50501 GCACACTTGT TTCCAATAAA GAAAAATCCC GACTTATAGA AAAAATTTTT 50551 TCCTATGATA CGACCTTCCC TGAGTTCAAG ATCCATAGTT TTGAGGGAAA 50601 GTCTCGAGCT TTTATTAAAG TTCAAGATGG CTGTAATTCT TTTTGCTCGT 50651 ACTGCATTAT TCCTTATTTG CGGGGGCGTT CGGTTTCTCG TCCTGCTGAG 50701 AAGATTTTAG CTGAAATCGC AGGGGTTGTA GACCAAGGAT ATCGCGAAGT 50751 TGTAATTGCA GGAATTAATG TTGGAGATTA TTGCGATGGA GAGCGTTCAT 50801 TAGCCTCTTT GATTGAACAG GTGGACCGGA TTCCTGGAAT TGAGAGGATT 50851 CGAATTTCCT CTATAGATCC TGATGATATC ACTGAAGATC TGCACCGTGC 50901 CATCACCTCA TCGCGTCACA CTTGTCCTTC GTCACACCTT GTTCTTCAAT 50951 CGGGGTCGAA TTCAATTTTA AAGAGAATGA ACCGGAAGTA TTCTCGCGGA 51001 GATTTTTAG ATTGTGTAGA GAAGTTCCGT GCTTCTGATC CTCGCTATGC 51051 CTTTACTACA GATGTGATTG TCGGATTTCC TGGAGAGAGT GATCAAGATT 51101 TTGAAGATAC TTTGAGAATT ATTGAAGATG TAGGCTTTAT TAAAGTGCAT 51151 AGTTTCCCTT TCAGTGCTCG TCGTCGTACT AAGGCATATA CTTTTGATAA 51201 TCAGATTCCC AATCAGGTGA TCTATGAGAG GAAGAAGTAT CTTGCTGAGG

51251 TTGCTAAGAG GGTAGGCCAG AAAGAGATGA TGAAGCGTTT AGGAGAGACT 51301 ACAGAGGTGC TTGTTGAGAA AGTAACGGGG CAGGTTGCTA CGGGTCACTC 51351 TCCTTATTTT GAAAAGGTTT CTTTCCCTGT TGTAGGAACG GTAGCTATCA 51401 ACACTCTAGT TTCTGTGCGT CTTGATAGGG TAGAGGAAGA AGGGCTGATT 51451 GGGGAGATTG TATGATAGAT ATAATGCAAC ATTTTAAGCC CTATACTATG 51501 GTCCCAGGAC AAAAACTCCC TATTCCTGGA TCTTTGTTAT ATGCTCAGGT 51551 ATTTCCTACC CTGTGGCGTC TATTTTCTTC GAAACACGAA ATCTTAAATG 51601 AGCAGACCTT ACAGGTGCAA GGGCCTTTAA AACGCTTTGC TGTTTTCCAA 51651 GATTTACATC GTGGGGGGCT TGCAGTGACT TCTGAGCGCT ACAAGTATTA 51701 TCTCCTTCCC TCGGGAGAGT GCACACAATC TATCAAAGGG AAACTGCCTT 51751 CGGCAGCGCA AGCAGGGCCC CTGTTATCTC TTGGGGTGCA TAAGCATGCA 51801 GATTGGCAAA AGGTCCGTTG TCGTCGTGAT CTTAAAGAAA TTCTTCCCCT 51851 ATGGTTCCGT TTCGCCGCTA TGGCTCCTAA GGGATCCTAT CGGGATCTAG 51901 AGACGACGGC TATCGGTAGC TTGGTAAAGA CTGCCCATCA AAGAGTTTTA 51951 CATAGGGAAA CTACAGAGAT TGCTCCTGCG TTACTCTCCA TAGCCCTTGC 52001 GGGATTTTCA GAGTGCTTTC TTCCTAGGAG CTATGATGAA GAGTTCCAAG 52051 GAATCCTCCC CCAAGATGGA GATCCAGAGG GGGGAGTTCC TTTTGAGCTT 52101 CTCTCGTATA GCTTTGGTAT GATCCAAGAT ATTTTTCTGA GACACCAGGG 52151 ACAGCTAGTA GAGATCCTTC CTGCATTACC TCCTGAATTT CCTTGTGGCC 52201 GCTTGATTCA TGTTGCCCTT CCTAATCTTG GGACTTTGTC TATCGTCTGG 52251 ACTAAGAAAA CTATCCGTCA GGTCGAGCTC CATGCAGAAT ATAGTGGCGA 52301 GGTATTTTTA AAGTTTTGTT CTTCACTATG CAGTGCGCGC CTTCGGGAAT 52351 GGTCGGAGCG ACGTCTCTCT GGATCTAAGA GACTTTCTTT AGGAGAAACT 52401 CTGGAGATAA AAGCAGGAAC CACATATTTA TGGGATTGTT TTCATAAATA 52451 GATAGCCTTC CATGGTTGAT AAACTGATCC ATCCTTGGGA TCTTGATCTG 52501 CTCGTCTCAG GACGACAGAA AGATCCCCAT AAACTCTTAG GGATCCTTGC 52551 TTCTGAAGAT TCTTCAGATC ATATTGTTAT TTTTCGTCCA GGGGCGCATA

52601 CGGTTGCTAT TGAACTTCTA GGAGAGCTTC ACCACGCTGT AGCTTATCGT 52651 TCGGGGCTCT TTTTCTTATC CGTTCCCAAA GGAATCGGAC ACGGGGATTA 52701 CCGTGTGTAT CATCAGAATG GACTTCTCGC TCATGATCCC TATGCGTTTC 52751 CTCCTCTGTG GGGAGAAATT GATTCTTTTT TATTCCATAG AGGAACGCAT 52801 TACCGCATTT ATGAACGCAT GGGGGCAATC CCTATGGAAG TTCAAGGAAT 52851 CTCAGGGGTG CTCTTTGTTC TTTGGGCTCC CCATGCGCAG AGAGTCTCTG 52901 TAGTCGGAGA TTTTAATTTT TGGCATGGCC TTGTCAATCC TCTACGTAAA 52951 ATTTCCGATC AGGGGATCTG GGAGCTTTTC GTCCCAGGCT TGGGAGAGGG 53001 AATACGGTAT AAGTGGGAAA TCGTTACCCA ATCGGGGAAT GTGATTGTAA 53051 AAACAGATCC TTATGGGAAG AGCTTTGATC CTCCACCCCA GGGTACAGCT 53101 CGTGTTGCGG ATTCTGAGAG CTACTCTTGG AGTGATCATC GTTGGATGGA 53151 GAGGCGCTCG AAGCAGAGTG AAGGGCCCGT CACGATCTAT GAAGTGCACT 53201 TAGGCTCTTG GCAATGGCAG GAGGGAAGGC CCTTAAGCTA CAGCGAAATG 53251 GCGCATCGCC TTGCTAGCTA TTGCAAGGAA ATGCACTACA CTCATGTGGA 53301 GCTTCTTCCC ATTACGGAGC ATCCCCTGAA TGAATCTTGG GGCTATCAAG 53351 TGACGGGATA TTATGCTCCA ACATCAAGAT ACGGGACTCT CCAGGAGTTT 53401 CAGTATTTTG TAGACTATCT ACATAAAGAA AATATTGGTA TTATTTTAGA 53451 TTGGGTGCCG GGACATTTTC CCGTAGATGC GTTTGCTCTT GCCTCTTTTG 53501 ATGGGGAGCC TCTCTACGAG TACACGGGGC ATAGTCAGGC TCTTCATCCC 53551 CACTGGAATA CGTTTACCTT TGACTACAGT CGTCATGAAG TGACCAACTT 53601 TTTACTAGGG AGTGCTTTAT TTTGGCTCGA TAAGATGCAT ATTGATGGCT 53651 TACGTGTGGA TGCTGTGGCC TCTATGCTGT ATCGTGATTA TGGCCGTGAA 53701 GATGGAGAAT GGACGCCTAA CATCTATGGA GGTAAGGAGA ACTTAGAGTC 53751 TATAGAATTT TTGAAACACT TAAATTCTGT AATTCATAAG GAGTTCTCTG 53801 GAGTGCTCAC CTTTGCAGAG GAATCCACAG CGTTTCCAGG AGTCACTAAG 53851 GACGTAGATC AGGGAGGTCT GGGGTTTGAT TACAAATGGA ACTTAGGTTG 53901 GATGCACGAT ACCTTTCATT ACTTTATGAA GGATCCCATG TATCGTAAAT

53951 ACCATCAGAA AGATCTGACA TTTAGCCTTT GGTATGCCTT CCAAGAGTCT 54001 TTTATTCTTC CTCTCTCGCA TGACGAGGTG GTCCACGGTA AGGGCAGCTT 54051 AGTGAATAAG CTTCCCGGGG ATACCTGGAC CCGATTTGCT CAAATGAGAG 54101 TGCTCTTGAG CTACCAGATC TGTTTGCCTG GGAAAAAGTT ACTGTTCATG 54151 GGTGGGGAAT TCGGACAATA CGGCGAGTGG TCTCCTGATC GTCCCTTAGA 54201 TTGGGAGCTT TTGAATCATC ACTACCACAA AACTTTGCGA AACTGTGTCT 54251 CTGCATTGAA TGCGTTGTAT ATTCACCAAC CCTATTTATG GATGCAAGAG 54301 AGCTCTCAAG AGTGCTTCCA TTGGGTAGAC TTCCATGATA TAGAAAACAA 54351 TGTCATTGCC TATTATAGAT TTGCAGGCAG CAATCGTTCT TCGGCGCTTC 54401 TCTGTGTCCA TCATTTCAGT GCGAGTACTT TTCCTTCCTA TGTTTTAAGG 54451 TGTGAAGGTG TAAAGCATTG TGAACTCCTT CTCAACACTG ATGATGAGTC 54501 TTTTGGAGGC TCAGGGAAGG GAAATCGGGC TCCTGTGGTC TGTCAAGACC 54551 AAGGGGTCGC TTGGGGTTTG GATATAGAGC TCCCTCCTTT AGCTACTGTG 54601 ATCTATTTAG TTACTTTTT CTAAAAATTT AAATACTTTA TTTGTAAATT 54651 GTTGTGGGAT TGTTCTATTT TGTGGTGTAG TTGATATTAA TAATTTATTT 54701 TATAATTAAA AATAATTATT AGTATTTCTT TTATGTCTAC ATCACCAATT 54751 AGCAACGATC CCCGATATTT GTCTTTGTCT AATGCAACTG AGAAAACTTC 54801 TCTTCTTGCA AATAGCCGGA GTCTCTCGCC AGTACCAAAT TCCCTAGTTC 54851 CTAGCAATCC TGAAGATACA GGATTGCGAA AAAGTATTTT CACCCATTCC 54901 GTGACTTTAT TTGCTGGCCT GGTTGTTTTG CTGGTAGCGG TTTCTGTTGT 54951 TGTTGTCGCT TTGACCGTCT TAGCTCCCGG AGTTCCTCAG GCTATTCTTC 55001 TTGGAATCGC CATTTCAGGC GTGGGTATTG GTGGATTTTC TATAATGAAG 55051 AGCTTGGTTT ATATGGTCCG AGACTATATG TCCCCCAGGA TGCAGGAGTC 55101 GAGCAGAATC AAAAGTGCTT TAGCTGTAGG GACTGGATTT ACTGTCATGG 55151 GTTTGGTCAT GÄAGGTGGGG GCGAATTTTG TTCCTGGAGG GTATGGGGGT 55201 CTCGTGGGTA GCTTGGGATC CAGTGCGTAT TCCCGGGGAA GCCAAACCAC 55251 ATTAGCAAGC TTCAGTCATT ATATTTATAC TAAGTTTTTC CGTTCTGAAA

55301 AAGTTGCTAA AGGGGAGAAG CTTACAGAAG CAGAAACTAT AAAAGAGGCG 55351 AAAAAATTAC ACTATATCAC GTTGTCAATT GCCACTATTG GCGTTGGTCT 55401 TGCGGTTTTG GGGATTCTCC TTGCCATTGC AGGAACGGTA TTGCTAGGAG 55451 GCGCTCCCGC AACGATTGCT ATTATTTTAG CTCCCCCTTT AATTTCTATA 55501 GGGCTTACGA CGGTTTTGCA AACGATACTC CATAGTAGTA TCGGAAAGTG 55551 GAGAGCCTTT CTGCTTACTC AAGAAAAAA AGATCTTTTT GTAGACACCT 55601 CCCTGAAAGA CATTCGCTTA GAAAAATTGC CCCCCAGTGA GGTGGAAGAG 55651 AGTGAAACTT CCCAATCTGT GATAGAAGTT CCAGATTCAG AGGGGATTGC 55701 AGAGACGAGG ATCTCTGCGG AAGAAATCGA TACGAGGCTT TCCCTGACGA 55751 CAAGACAGAA GGTCATCTTT GCTCTTGCGA CACTCTTGCT CTTAGCAAGT 55801 ATTGCTGCCT TCATAGTCAC GGGATTTGGT GGATTGACAG TCATGCAAGT 55851 TCTCCTTGTT GCTTCTGTAG GATCGGCGGT TGCTTCTGTA ACACTCCCTA 55901 TGGTTTCCTC AGGATTTTCC TACGTCGCCT ACCAACTGAA AGCAAGATTG 55951 AATATCAGTA AATTACGTTG GAAAGAAGCA AAAAATAAAA AGCGGGTGCG 56001 CCAGTTCTTA ATTGAGTCTG GAGTGATTGC CTCGGATCGA GAATTTAACC 56051 AAATGTGGAA GACAGTCTAC AAAAAACAGA TTCAGAAGAC TGACGCTGCA 56101 ATTCGTGAAG AGGTTCGCAA TTTTGAGAAG GGTGGGGAAG TGAACAGCGC 56151 CCTTGTTGGT GGAATCTTAC TTGGTGTAGG AACTGGGATC ATGCTTCTTG 56201 CCCTGGTCCC TGCATTTGCT CCTATCGTTC CTGGTATTCT TGCTCTTGGA 56251 GGATCGACGT TAGGAATCGC GGGATCGATT TTAATGAGGA AGTTTGTCAA 56301 CTGGCTCTAT GATGAGCTTG TGAAGCTCTA TGAGCGTCGA CGTAATCGCC 56351 GTGAGCTTCT CTATGGTCCT GAAAGTAAAA TGCGCTCCAT TGCTACGGAT 56401 TTAGTTGTTG AGGCTCTTGC TGCTAGCCAC GATCATCTAT TTGATCTTGA 56451 TGGTCCCGTA GATTTTATTG ATGTGGATGT AGATATAGAT GGAGCTGCTT 56501 AGGCCAGGTC CTTGAATGTA AGATCCTCGA GCTTTGGGAG CTTGTCTGCC 56551 TTCTCTAATT TTATTTTCTC TTTTACATCT AGTAGTTTAC TTAATTATAT 56601 AATCAGCATT CTTTTTGTTT ATTTTAATTT ATATTTTGTT TTTAAAATAT

56651 TTTTTATTTT AACATTTGTT TAATAGTTTT TATTAAATAA TTTATCATTT 56701 TAAGGTTCAA TTATGGCAGT TGGTGGCGTA GGCGGCTCAA GATCTCCTTC 56751 CCCCATTCCT CCTAATAGAA GGAATAGTGA GGATGGAAAA GTAAGTCCTA 56801 AAGACAACTT AGGGGAACAT ACAGTTAGCA GTAGTGACAG TAGTCTTGCA 56851 AGTCAGGGCC CTACAATAGA AGAGAGAAAA GCCCAGTTAG GCGGGACTGA 56901 TAAAATTCCT TTGCCATCTG TCAAAGAACC CGGAGATTCT CAAACTTCAG 56951 GACGTTCTGG GGTACTTCAG AGAATTTGGA AAGGCGTTAA AGGGGTCTTT 57001 AAAAAAACCC CTCAAGCGCG TCCTGAAGTT TCTAGTCCAC GTCTTCCATC 57051 CCATGTGCAA CATGGCCAAC GTCTTCCTGG ACTCGAGGGC TTTAGAGATC 57101 GTATCCAGAA AAGATCTGAA AATCCAGAGG CAGATTTAGG GAAGATGAAA 57151 CGTTCCTATT CTGATGGTGA CCTTGATCGA GTAGGACACG ATTCTAATGA 57201 AGATTCTACA GAGGATAGCC GTTCTGAAGG AGGAGAGCCT TCTTCAAAGA 57251 GTTCTTCCTT CTTATCAGGA GTTCGAGGAG CGGTGTCTAA AGTTCATGGT 57301 GCCCTAGGTG ATATTAAAGG AAAGTTCCAG CGTTCTGCTT CCGAAGATGA 57351 TTTAACAACT CAGGGCGAAG ATTCTGCCGG CGATACTGTA AAAGAAAGGC 57401 GTTCCGAAGA AGCAGAGGCT TCTTCGAAGA GTTCTTCTTT TTTATCAGGA 57451 GTTCGAGGAG CGACGTCTAC AGTTCAGGGA GCCTTAGGTG ACGCTAAAGA 57501 GAAGGTTTCG GCGTTCGGAG AGCAGGCTGC AGGTGCAATC AGATCAGCAC 57551 CAGGGAATAT CAGAACTAGA TTCCAACGTT CTTCATCGGA AGGTGATCTT 57601 TCTAATGTGA ATAAAGCAGC AAAACATCTG CGTAAGGCTT TAGAAAATTT 57651 GGAAAAAGTA GCTCCAGAAC AAGTGTCACC AGAGGTGGCT TCTAGGGTGC 57701 AATCTCTTCT TGCACGCATG GAGCAATTGA CTCATCAGGA ACCTCCTACT 57751 GTGGAGGATC TTATTACTTT CGTAGAATCC AATGTAGGTA GTGATTCTGT 57801 GGAGTATGCA TCCATCGTAC CTCAAGATGG ATCGCAAGCC CCAGCAGAGA 57851 CTGCGGAAGC TCCCGAAACA GGTGGGGTAG AGGGATCTGC AGCGCAGGGA 57901 GCATGGAAAG CGTTACGGGA TTTTGTAGTT AGCATATTCC AAGCGGTAGC 57951 GAGCTTCTTT AGGGCAATTG CTTCAAGATT AAGTTCAGCA CGACGTGAAT

58001 CAGCTGTAGA TGATCTTGCA TCAGAAAGTA ATACACAATG GTTTGTGGAG 58051 CAAGAGGGCG TTTCAAATCC ATCGGCTGCA CCTAGCTTAT CTTTTGCGGA 58101 AGAGATCGCT CGTAGAGCTG CAGAAATGAG TAACAGAAAT GCCCAGAGTC 58151 TTGAAAAATT GGAATCAGGC AATGTGACTG ATCCTGTCAT TCAACAAGGC 58201 TTAGGATTAG CTAGATCATT TGCTCCAGAG GGACAGTAGT CGTTATCTCA 58251 CTGTTTCTCT ATGCGCAAGG GAAACTTGAA GAGTTTTAAT TAAAACTCTT 58301 CAATATGTTG ATTATTTTAA TATATTTAAA AGCATTTTTG TTGTTTTTTA 58351 ATAAAATTAA ATTGTTTCAG AAAAAAGATT ATTCTTTTTA GGAAGTGTTT 58401 ATGGCATCAG GAATCGGAGG ATCTAGTGGA TTAGGAAAGA TTCCACCTAA 58451 AGATAATGGG GATAGAAGTC GATCGCCCTC TCCTAAGGGA GAACTTGGCA 58501 GCCACGAGAT TTCCCTGCCT CCTCAAGAAC ATGGAGAGGA AGGAGCTTCA 58551 GGATCTTCGC ATATACATAG CAGTTCCTCT TTTCTACCAG AAGATCAGGA 58601 GTCTCAGAGC TCTTCTTCGG CAGCTTCTAG CCCGGGATTT TTTTCTCGCG 58651 TACGTTCTGG GGTAGACAGG GCCTTAAAAT CATTTGGCAA CTTTTTTTCC 58701 GCAGAGTCTA CGAGTCAAGC GCGTGAAACG CGACAAGCTT TTGTTAGATT 58751 ATCAAAAACC ATCACCGCGG ATGAGAGACG GGATGTCGAT TCATCAAGTG 58801 CTGCTGCTAC AGAAGCCCGA GTGGCAGAGG ACGCGAGTGT TTCAGGCGAA 58851 AATCCTTCTC AGGGGGTTCC AGAAACCTCT TCTGGACCAG AACCTCAGCG 58901 TTTATTTCT CTTCCTTCAG TAAAAAAACA GAGCGGTTTG GGTCGGTTGG 58951 TACAGACAGT TCGCGATCGC ATAGTACTTC CTAGTGGGGC TCCACCTACA 59001 GACAGCGAGC CTTTAAGTCT CTACGAGCTA AACCTCCGTT TGAGTAGTTT 59051 ACGTCAGGAG CTCTCTGACA TACAAAGTAA TGATCAGTTG ACTCCAGAGG 59101 AAAAAGCAGA AGCCACAGTT ACCATACAAC AGCTGATCCA AATTACAGAA 59151 TTCCAATGCG GCTATATGGA GGCAACACAA TCTTCGGTAT CTCTAGCAGA 59201 AGCTCGTTTT AAGGGGGTAG AAACTAGTGA TGAGATCAAT TCCCTCTGTT 59251 CAGAACTGAC AGATCCTGAG CTTCAAGAAC TCATGAGTGA TGGAGACTCT 59301 CTTCAAAACC TATTAGATGA GACTGCCGAC GATTTAGAAG CTGCTTTGTC

59351 CCATACTCGA TTGAGTTTTT CTTTAGACGA TAATCCAACT CCGATAGACA 59401 ATAATCCAAC TCTGATTTCT CAAGAAGAGC CTATTTATGA GGAAATCGGA 59451 GGAGCTGCAG ATCCTCAAAG AACTCGGGAA AACTGGTCTA CAAGATTATG 59501 GAATCAGATT CGCGAGGCTC TGGTTTCTCT TTTAGGAATG ATTTTAAGCA 59551 TTCTAGGGTC CATCTTGCAC AGGTTGCGTA TTGCTCGTCA TGCAGCTGCT 59601 GAAGCAGTGG GTCGTTGTTG CACGTGCCGA GGAGAAGAGT GTACTTCTTC 59651 TGAAGAGGAC TCGATGTCGG TGGGGTCTCC TTCAGAAATT GATGAAACTG 59701 AAAGAACGGG CTCTCCGCAT GACGTTCCAC GCAGAAATGG AAGTCCACGT 59751 GAAGATTCTC CATTGATGAA TGCCTTAGTA GGATGGGCAC ATAAGCACGG 59801 TGCTAAAACC AAGGAGAGTT CAGAATCAAG TACCCCGGAA ATTTCGATTT 59851 CTGCTCCCAT AGTGAGAGGT TGGAGTCAAG ACAGTTCCGT CAGTTTTATT 59901 GTTATGGAAG ATGATCATAT TTTCTATGAT GTTCCTCGTA GAAAAGATGG 59951 AATCTATGAC GTTCCTAGTT CCCCTAGATG GAGTCCTGCG CGAGAGTTGG 60001 AAGAGGATGT TTTTGGAGAT TATGAAGTTC CTATAACCTC TGCTGAACCA 60051 TCTAAAGACA AGAACATCTA CATGACACCT AGATTAGCAA CTCCTGCTAT 60101 CTATGATCTT CCTTCACGTC CAGGATCGTC TGGAAGCTCA CGTTCTCCGT 60151 CTTCAGATCG CGTACGAAGC AGCTCACCAA ATAGACGGGG TGTGCCTCTT 60201 CCTCCAGTTC CTTCACCTGC TATGAGTGAG GAGGGGAGCA TTTATGAGGA 60251 TATGAGCGGT GCTTCAGGTG CAGGTGAAAG TGATTATGAA GATATGAGCC 60301 GTTCCCCCTC TCCTAGAGGC GACTTGGATG AACCCATATA TGCTAATACT 60351 CCTGAAGATA ATCCATTTAC TCAGAGAAAT ATAGATAGAA TTTTACAGGA 60401 GAGGTCAGGC GGTGCTTCCG CTTCTCCTGT AGAGCCTATT TATGATGAGA 60451 TCCCATGGAT TCATGGCAGG CCCCTGCTA CACTTCCAAG ACCCGAGAAT 60501 ACATTGACTA ATGTTTCGCT TAGAGTGAGC CCAGGGTTTG GACCAGAAGT 60551 AAGAGCCGCT TTGCTTAGCG AGAGCGTGAG TGCTGTTATG GTCGAAGCAG 60601 AGAGTATTGT TCCTCCAACA GAGCCGGGGG ACGGAGAATC AGAATATCTA 60651 GAGCCCTTAG GGGGACTTGT AGCTACAACG AAAATCTTAC TACAAAAAGG

60701 ATGCCTCGT GGAGAGTCGA ATGCTTAGGA TTTAAGTAGT TCTTTCGAAT 60751 CTCTAGTGAG GATGTATCGG GTTCTTAATT TTTATGGGGG AAACGTATCT 60801 GTGGATTTCC CTTAGCTTCT CCCATAAGAT TCATGATGGT AGAGAGTGTT 60851 CACTACTCTC TATTTGGTCT TTACAGGTTG CATTGTCTAT ATAACATGCT 60901 TTTAAAACTT AAAGGTCGTT CCTGCGTGAA GGTAATGTGT CGTCGTTGAT 60951 GAGGATACCG AGCCTTGATA GTCTAAGAAC ACCGAAAGTT TAGGGAAGAT 61001 AAAAATTTGG TTTCTTCCTT TAAAAGCAAT GGCATTGCGA GCAAGGGTGG 61051 TTCCTGATAG GAGCCATGAC GATCCACTAG ATTCTAGACT CACGTTGATC 61101 TCAGGATTTT GTTGGTAGAG GACAGGCTGA TAAGCAAGCT CTATGTTCCA 61151 ATAGGTAGGA AGACGGAACT TGGATTCCCA AGCGCTCTGA ATTCCCAGAG 61201 GGACTGTCAG GTTATATAAG GGTTTATGAA CAGAAAATTT TCTAGCTTTA 61251 TCTCCACTTT CTTGAAACGC AGTTTGATTA GAACGAACGG CAATTGCTTG 61301 GATAAAAGGA GTGAAGTGGA GAGGTCGTGA TCGCCATTGT AGAGATAGAG 61351 AGCAAGAGA AGCCGCCCCT AATGTCGTAC TATAACATTT GCCTTCCGTT 61401 TGTATTTTC CAGAATATCC AGATGCTTTG ATATGGTGGT TGCTGTAGCT 61451 GTAGGCTAGA GATGCAGATG TAGAGAATCT CTCTTGCAGC CAAGGATTAT 61501 TGATCTGGAG CGCTACAGTT GTCGTATGCG AAGCCACGGA ATTGTCGGAG 61551 TGGCTCTCGT AGAGATTACT GAAAAGTTGG GAGAAGTTTA CACCAAAGCT 61601 ATGATTAGAA GCAGTGTTTG AGGTTGTTCC CAAAGAATAA CCCGTAGCTT 61651 CCATATGGAA TCCTTTCGCA TCATTGTTGC TATTTTGATG CACGAAGAGT 61701 CGAGTAGCTT CTCCAGAAGC TGTAGGTGCT ATTTGGCCTT GCTGTGTTTG 61751 ATAACGTAGT GTCGCAAATA AGTTATGGAA AGATTGCCAG AAGGCAGATA 61801 GGGCAATGTC TCCTTTGTTT TCTGGGTTTA CCTTATATCC TGTAGGTGTC 61851 CAATCACCAT AAAGCTGGCG ATGTAAAGTA TTCACAGTAT CTTCAGAAGA 61901 GGTATCAGAA GTTGTGATTG TTTCGATCCA GTAAGGGGAC CAAACGCCTT 61951 GGTAGCCGTA GTGTTGAGTT GTATTTAGAC CCTCAGGGTA GAAATTATCC 62001 GTATTAATAT GTTTAGCTGT GACGTCTAAG AGATACAGAA GAGGAACTTC

62051 TGCGATAGGT TGGGCAAGGT CTGCAGTATC ATAGGGATCT AGGTTCTCGT 62101 CATCCAGTAG GCTCAAAGGT CCTGAGAGAT TGATTATAGG GTTATTATCT 62151 TCGCTATAGG GTGCTGATGA ACCTGTGGGG CGAATCCATA GCTTGGGAGC 62201 AACTCTGTTG CCTAAGATAG AGGGAAGGTT AATTGCAAGA TTATTGATGT 62251 TAATTACAGA ACCCACACTA CTGCTACTTT GTTCTTCGTC TGTTGTAGAA 62301 AACACAGCTC TACTGCCTAA CCGTAGAGTC CCACCAAATT GATCAAATTT 62351 ATAGACTTTC CACTCTGCTC GATCTTCAAG AGCGAGTGTG CCGTTGTACA 62401 GTCCAATGTG GTTTCTGAAA TGTGAAATGA AGTCATCACG AGAAGTCGAT 62451 GTATCCGGAA TATATGTTGA GGAGAACAAG ATAGTTCCGA GGTGTTCTGG 62501 ATTAGGATTA AATTTTTGGA TAGAGTTTTG TATAGTATAT CTTTGTAGTA 62551 TGGGATCATA GAAGGTAGCA GAATGACCTT GACTTGCTCC AACTGTTAAT 62601 GAGACATTAC GCGTGCAGTT TACAGAAACA TGATTGCTGA AAGTATCTTT 62651 GAAGTGTCTA TTATTATAAA AAATAATATC TCCCTGATCA GCAAATAAAG 62701 TGCATGCACC ATCTTGACGG AGCATGATAG CGCCGCCCCA AGTTCCCTGA 62751 TTGTTTGTGA AATAGACGGG ACCACTGTCT TGTATAGTTA GAGATTGTGT 62801 ACAGATAGCA CCTCCATCTC GTGCTGCAGT ATTATTATCG AAGGCTGCAA 62851 TTCCTGGGTT GTCTTTTATA GAACAACTAA TGCAATAGAT AGCCCCTCCA 62901 GAGGAATGGT TAGCAGAGAT GTCCGCTTCC ATGGCAAAAT TATTGTTGAA 62951 GATCACAGAA CCGGTATTCT TTGTAAGAAT GCACTCTTGA TGTACTCTTA 63001 TTGCACCACC CAGACCTGAT TGGTTATTCA AAAAATAGAT AGGCTGAGAA 63051 TTATTCTCAA TTCTACAAGC ATTAGCGAAC AACGCGCCCC CCGCTGTTCC 63101 GCCTGCAGCA TTATTAAAAA ACAGGCAAGG GCCAGTGTTG TCCTTAATGT 63151 TTATGATTGC AGCTTGGATT GCTCCTCCTG AAGATTTTGC CTTGTTGTTA 63201 ATGAAGTATG CGGTTCCTTG ATTTTTTGAG ATTGTAACAT TTTTCGAACA 63251 TAAAACAGCT CCCCCTGTAC AAGTATCAGC GAAATTACTT GCATTAGGAA 63301 AGCTTAAATT CCCAGAGAAA ATGATGGAAC CATGATTCTC AGAAAGATCG 63351 AAATTACCAT TCACATACAT CGCACCAGCT CTTTTAATAG CAAAGCTATT

63401 TAGGAAAAGA ATTTGGTTTT TTGTATTCGT TATGGCAAGT GATTTGCAAG 63451 AGAGAGCACC GCCGTCTTGA GAGAAGTTTT CGAACCAGCT TTCTATGGAA 63501 TTCTGGTGAT CGAGGACAAT GTCTTGGTTA GTGTCATCCC TAACTCCAAA 63551 AAGTGTTGCT CTATGAGAGT AGGGAGTCAT GTTAGTAAGA GTATCAATTA 63601 GAGGGAAGAG TGTTGTGAGT TGATTTGCTT GATTATCAAA ATAGTCAGAC 63651 AACGGAGTCG CATTAAGGAG TATTGTAGTT TTACCTAAAA TTAAGGCTCC 63701 AACAAAGAAG GAAGATTTAC TAAGGGATCT GTTATTTTGC ACTGTTTATT 63751 CCTTGATGTT TTCTTTGTGG TTAAAATGTG CAATGACTCT CAGCTTTAAG 63801 GTATTGACCT ACAGTTGACG AGGAGACTTG AGCTGAATAA TCTAAGGATA 63851 AGGTTACTCT TGAGAAAAGT TGGGAAGTAT TTTTTATTTT TGCAGCTACG 63901 GAATTATAGG AGACGGGAGT TGCTTGTGTT GTCCATGTTC CATTGCTGAT 63951 GAGTAGTGTC GTGAACATTT CTGGATTTTT TCTGTATAGG GTAGGTACGT 64001 AGGATATTTC CGTAGTCCAT AGCATGGGGA TATGATGTGA AGTTTTCCAT 64051 TCAGAACGGA AGCCTATGGG AGAGGAAAGA TCTGTAAGGG GATGTTTTGG 64101 ATGGAATTTT CTTATATGGT CTCCAGTTTC TTGGAACGAG GCCTGGGAAC 64151 AGCGCAGAGC AATGGCACTG ATAAAGGGCT GGAGTTCGAG AGTGCGGGTG 64201 ATTCTAGCTG GTAAGAATGT GCAGTCTAGA GAGGCTACCA AAGTGTGGTT 64251 ATTAAAGAAG GCTTTGGACG ACCCTTTTAA GATTTCTGTA TAGTGGCAAA 64301 GCATATGGTG ATCTCCGTAG CTATAACCTA GGGATAGCCC TGTAGAGATG 64351 AAGTCCCTGA AGAGGAGACT GTCGAAGCGG AGTCCTGCAA AGTAGTTGTG 64401 GGAGGAAGTC GTACTTGGAG ATTGACGTTC TCTAGTTTTG GAGAACATTT 64451 GTGCGAATCC TAAAGAGAAA CTATGTCGTG CTGCAGTTTT TGCTGAGGTT 64501 GTTGCTGCAT AGCCCGTAGT ATGGTTTCGG AAGCCTTTGC GTCCCTCGCG 64551 ATTATGTTGG TTAATTAGAA GCCCGAGTCC TTGCAGAGAG GCTTCAAGGT 64601 CATGCTCTTT GAGGTTTTGT GGAGGTAAGA TGCGGATTCC TAACAGAGCG 64651 TTATAGGCAG ACTGCCATAA GGTATTAGCA ATAAATTCTC CGTGACGTTC 64701 CGGGTTAGGG CGGTATCCTA CAGGAGTCCA GTCTACGTAG AGCTGCCTGT

64751 GGTTTGTATT GGTCTGTTCC GGTACTGTAG AGCTTGTTGT AGTCGTAGTT TCCATCCAAT AGGGAGACCA GATTCCCTGA TATCCATAGT GCTCATCTAA 64851 GTTCATGGCT TCTACAATGA GATTCGAAGT ATCGATTTT TTTGCAGTCA 64901 CATCGAGGAG GTAGAGGAGG GGGGATATCC TTTCGAGGTT CAGAGAGATC 64951 TAAGCTATCA TAGGGGTTTT CATTTTCATC GTTTAGAAAA GTCAAGGGTC 65001 CTGAGAGAGT GATAGTAGAA GAAGTGTCTT CAGAATAGGT GGATCCTGTT 65051 AATGTAGGAT AAATCCAGAA CTTTGGAGCT GAGGCTTCTG ATTGTAAAAT 65101 AGAAGGAAGA TTGATCGCGA TTGCATTAAA ATTTATGGAG CTTCCCGGGC 65151 CTTTCGTCCT GATTAATGCT GCGTTTCCTA AACGTAGAAT GCCCCCAGTT 65201 TGCGATAGGG TTTTGCAAGA AATAGCAGCC CGATCTTCAA TAGCGAGCAC 65251 ACCCCTTTCA AGTCGTGAAG AGTTAGAAAA TTTTGATAGG AAGTTCAATG 65301 GATTTGTTGC GTTAGAATCT ACATTGATTC CGGAAAACAA CACGGTGCCA 65351 AGGTGATGGG GTTCATAATT AAATACTATA GGATCTGTTG TCGTCTGATC 65401 GTGATCTATA GGATCATAAA AGAGAATTTT ATAACCCTGT CTTGCTCCTA 65451 GTTTTAAGTT AATCCCCGGA GCAGCATAGA GTGCATTTCT ATATCCGGGT 65501 TGAGGAGAAG AAGATGTGAT TGTATTATTG TTAAATAGAA TATCGCCGTA 65551 GTCTGCAGAG AGGAAGAAT TTTGAGGAGT ACTTCCTATA CCAGAAAGAT 65601 TGATGAGAGC CCCTCCTGAA GTCGCAGAGT TATTAATAAA TGCTGTCGGA 65651 CCGTTATTTT GGAAGATGAA AGATCTCGTG TGTATAGCTC CGCCGCTAAG 65701 TGCTGCCGTT TTATTGTTAA AGATAAGACC TTTGGGATTG TTCTCAATGA 65751 CTAAGGAGGT ACACATGATA CCGCCACCAC CGGGATATAG TTTTCCTGAT 65801 GCTGTGTTAA TTGAGGATGC GGAGTGATTG CTGATCTCTA TAATTTCTTT 65851 ATTTGAAGAG ATAATGACTC CTAGGGCAGA AAATATACCA CCCCTTGTCC 65901 TGAAGCATTA TCATTGATTT GGATGTTTTG ATAATTCCTT TCTATATTTA 65951 CGGCTCTGCA GAAGATGCCT CCTCCAAAGC TGGAATCTTC TAATGTTTGA 66001 TTTTTCTTGA TAACGATAGG ACCTAAGTTG TCTGTGATTG ATAAACTCAC 66051 CCCAACATAG ATAGCTCCTC CTCTATTTTT AGAGACATTA TCTAAGAATT

66101 GTCCTTCTCC TAAGTTATTT GTAATATAGC TATCTAGTGC ATGTATAGCC 66151 CCACCAAATC CTGAAGTTGT AGTAGTAGCT AAGGAAGCCG CATTTGAAAT 66201 AAACGAGTAG TTCTGATTCT TAGAAATCCA GCATTCCCGA ACACTGTAGA 66251 GAGCCCCTCC AGAACTGTGG GAGCTATTCC TTTCAAAACT TAAGTTTCCT 66301 TTATTTCAG AAATGAACAA ATTTTTACAT GCAAGAATGC CTCCATCCTC 66351 GTAGTGGTAG TTATAATCTA TAACAGAATT GATAGAGTTT CCTGTAATTG 66401 TGATGTCCTG GGTTATATCG TGACCGAATC CATTAAGAAG ATTAGAGGGA 66451 CGAAAGGAGG AAACCAAGGG CTCTGGAGTT ACATTCAGAC GGAAGCTTGG 66501 AGACGTGTGG AAAGCAGAGA TGTCTTTTCT AGACATCTGA CAAGAGGCGA 66551 GGTTAGGGAC TTCATTTCCT GATAAGGAAC AACATAGCGC AGTGGATAAA 66601 ATGCTGAGAC AAATGGGTCG CATAAACATC CAAACTTTGA TATGAAAATA 66651 GAGTTTGTCA TAAAAAATCA TCAATTTGTT GTCTAGAAGT TTATAAATTT 66701 ATATTTAAAT ATGCTTATTA ACATGTTGAA TATTACAGAG TTAGTAAACT 66751 CATATATTG AATGATTTTC TATATTTAAA TTTTGAGGGA GTCCTCTACC 66801 GATTGGGCTA ACGTAGCTCA TCGGATAGTT GTTAATTCTA AAGATTTCAA 66851 TTTTATCGTT CTTTTATTTT AGAATTTTAA GGTACTTCCT GCTTGGAGAT 66901 GGTGCGTAGA TGTCGAGGAG GAGACCGATC CTTGGTAATC CAAGAATAGA 66951 TCGAGAGAC GGAAGAGCC AGTTTGATTG TGGACTTTGT ACCCTAAAGC 67001 ATTGCGAACA TAGTTATGGC CTAGGATATC CCAGGAACCT CCGCTCGCAA 67051 GTAGCGTGAC ACCGATTTGG GGATTTTGTT GATAGAGTAC CGGTTGGTAA 67101 GAAAGTTCTA GAGTCCATTC TGTAGGTACG TGGAATTTTG ACTGCCATTT 67151 TCCTTGGATT CCTAGAGGTA AGGTCAGATT ATAGAAAGGC TTTTGAGAGA 67201 CAAACTTTCG GGGATTGTCA CCAATCTCTT CGAACGCTGT TTGGTGAGAA 67251 CGTATTGCAA TTGCCTGAAC GAACGGGCTG AGGTGAAGAT AGGATTTCTG 67301 TTGCCAAGGG AAAGAACAGC CGATAGCTGC TGCTAATGTA TGGCTATAAC 67351 ACGTCCCTTC TGCCTGTTCT TGATGTGAGG GATGTAGGCT GTGGAGGTGA 67401 TGGTCCCCAT AGCCATACGC TAACACTGTG GATGTTGCAA AGGCCTCTTG

67451 GAACCACGGA AGCTCAACAT AAAGTGAAGA GACTGTATTG TGAGCCGAGA 67501 CGTTGTTGCT TGATCCGATT TCTTTAGTGC GGGTGAAGAA CTGTGCAAAA 67551 CCTAAGGAGA TTTTCTGATG TAAAGAAGTT TCGGAGGATG CTTGTAAGGA 67601 ATACCCTGTA GATTGGATAC GGAATCCTGG AGCCCCGGGG ATGCTATTTT 67651 GATGAACAAA GAGGCCGTCG GCAATCCCTT GAATTTCTAA GAAAGGCCTC 67701 TCGATATCAG AATCACCAGT TCGATTATAA CTTCTTAATA GAGAGAACAT 67751 AGTATGAAAG GATTGCCATA GGGGAGTCGT AGCAAGATCT CCTTGGTATT 67801 CAGGATTGAC CTTATATCCT AAGGGAGTCC AATTGGCATA CAGAGCTCTG 67851 TAGAGGGTGT TTGCCGTCTC TATAGAAGCG TTATTTGTTG TTGTTATCGT 67901 CTCTACCCAA TAAGGAGACC AGATGCCTTG ATAACCGTAA TGCTCAGTCG 67951 CATTTAAGCT TTCAGGATGA AAGTTATCGG TATTGATATG ACGTGCTGTT 68001 ACATCCGATA AAGAAAGAAG ATGAATGTTT TGTAAAGGCT CAGAGAGATC 68051 TATACTGTCG TAGGGATCGC GGTTTTCCTC ATTTAAGAGT GTCAGAGGAC 68101 CTGATAAAGT AATTGTAGGG TTATTGTCCT CTGTGAAAGG AGCACTAGAT 68151 TGTAGAGGAC GGATCCACAA GGTAGGAGCT TTTCCTTTTG CTAAGATCGA 68201 GGGGAGGTTA ATCGCAAGGT TATTAATGAT GACCTGGGAG CCTACACTAG 68251 TTGATGGAGT CTCAGAGTTG GCAGTTGTTG CAATACTCGC CGCATGCCCT 68301 AATTTAAGGA TACCTCCTTT TTGAGTGAAC TTATAGAATT GCCATCCCGC 68351 ACGATCCTCG ATAGAGAGGA CACCATTGCG AAGTTCAGAG GTATTTTTCG 68401 AGCTGCTAAT GAAATTATTT TCGTAGTCAG AAGCTTCTGG GATATAGGCT 68451 GAAGAAATA AGATCGTTCC CTGATGGTTC GCATTGGGAT TAAAGATTAG 68501 AGGATTTGTA GTTGGATGTT GGTGTTCTAT AGGATCAAAA AAAGCAGTCG 68551 TATACCCCTT ATTAGCTCCA AGTTGTAAGT TGCTATTTGG TGTACAATGT 68601 ATGGCGTTGT ATCTACCAAA TGTGGTGAGG AAAACCTCAT TATTTTGAAA 68651 TGCGATATTT CCTTGTTCCG CGAAGAGTAG GCAGGTGCTG TCCTGTAGGA 68701 GCATAAGAGC ACCTCCCCAG TTTCCTTGAT TGTTGGTGAA ATATACGTGG 68751 CCACTATTTT TGATTGTCAA AAATTGTGTA CAGATAGCTC CGCCATCGCG

68801 AATGCAGTAG TTATTATTGA AAAGAATAGT TCCAGGGTTA TCGTCTATGG 68851 ATAGGTTTGT TGTATAAATC GCCCCTCCTG AACCATTTCC TGAATTTATC 68901 GAACCAGATA ACGCTGTGTT GTTATTGAAA ATCACCGACC CGGAGTTATT 68951 TTTTATCGCA ACAGTAACGC TTGTTTGAAT GGCCCCGCCA TTGTTCCCAC 69001 AGTTGTTCTT AAAATAAATA GGACGCGTGT TATCAGAGAT CGTTGTATTT 69051 TCACTACGAA GCGCACCCCC TCCACTAGGG GCTGTATTGT TAAAAAAGAG 69101 TAGAGGTGCC CTGTTGCTTT GGATGCGGCA GTGTCCATTG GTGGAGAGGG 69151 CTCCTCCCCA GTTGTTGACG GAATTGTTGA CAAAGTAGAA AGTCCCTTGA 69201 TTTTGAGAAA TCGTGAAGTC TCCATTACAG GCAATCGCAC CCCCACGAGT 69251 TTCTCCTCCT GTACTCGCAT TGTTAAGACC TCGATTGCTG AAAAAAATAA 69301 GGGGTCCTCT ATTCTTCGTG ATTGTGCAGG CTCCCTGGCA AGCAATCGCG 69351 CCTCCAGTCC CAATCGCGAG ATTTTTACTG AAGAAGGCAT GGTCTTCAAC 69401 ATTTGATAAT AAGAAATTAT TACAGGACAC AGCTCCCCCA GCCGATGTCC 69451 AAAGAAGAAG GATGTTATCA ATAGACTTGT AGTTAGAAAG TACAATGTCT 69501 TGAGAGGAAT TATGTCTATT TCCAACAAAC GTAGTTATTG GAGAAAATCC 69551 TGTAAGAGTG GAGAGAGAGT CTAAGAGAGG AAAGCTCGTA CGAAACTCTT 69601 CATCCCTCTC TAAAGCAAAC TTTTCAAGGG AGTCCGTTTG TAAACTATAC 69651 ACTGCAGGAG TCATCCCGAA CATGCAGGCT GTGAAATTCC CGAGATAGAA 69701 TAAAAACTTA GGAGGAGTCT TTGACACTAG AGGTTCCTTA ATCTTTCTGT 69751 TTTGGTCACT TATTGTTAAA ATCTCATTCT ACTCGCCACG TTTAAGTAGT 69801 GACTCAGCGT GGAGGAAGAA ATATCCGCAG AGTAATCTAA GGAGAGAGTG 69851 ACTTTAGGAA ACACCTGCAT GGTATTTTTC ACTTTGATCC CTAAAGCATT 69901 GTAGGTCACA GGAGTGGCCT GCGTCGTCCA CGTACCTTGG CTAATCAGTA 69951 ATTTCGAGTG GAGTTCAGGA TCTTGCCTAT AGAGAGTAGA GCGATAGGAA 70001 ATTTCTGTGA GCCAGACTAG GGGAACTCGG TGGTGGTTCT TCCAAGAAGC 70051 GCGGATTCCT ACAGGGAGGG AGACGTCCGT TAGGGGGCGG TGTAGGGAAA 70101 ATTCCCGAGC ATGGTCTCCA GATTCTTGAA ACGCAGCAAG ATTTCCTCGG

70151 ATGGCTAAGG CAGTAATAAA GGGATAGATC TGCAGGGACT CGCCGTGAGG 70201 TTGAGGTAAG AAAACACAGG AGAGAGCCCC TGCTAAGGTA TGGTTGTGGA 70251 AAGATCCCTG AGAGTTCCCT TCCAGGAGAC CCTGATACAT TGTATGGGTA 70301 TGTTCCGAGG TAAACATATA AGCAAGAGAC ACAGATAGAC GTATCCACTC 70351 TTTGAAGAGA GTATTTTCTA TGCACATTCC AGAGAAATAG TGGTGAGAGG 70401 ACGTGCTATT TTGAGATTCA TGTTCTTTAG CTTTGGAGAA GAACTGAGCA 70451 AATCCTAAAG AGAAATTCGG ACTTTGAGAA GAGGTTGCTT CGGTGGTAGC 70501 ACTATAACCT GTCATATGAC TACGAAATCC CTTAAAACCG TTTTTGTCTT 70551 TTTGATGAAC CAGAAGACCA ATGCCTTGTA GGGAAGCTGC ATGACCCTTC 70601 TCTTCATCCC AGGAGGAGAG GGAGTGGAGT CCTGCAAGAG CCGTATATGC 70651 CGATTGCCAC AAGGCATTCG TAATGAATTC TCCTCGACGT TCGGGATGAG 70701 GACGGTAGCC TAGAGGAGAC CAGTTTGCAT AGAGCAGCTT GTGTTTTGTA 70751 TTCGCGCCTA GTAGAGATGT AGGGTTCGTG ATTGTTGTAG TTTCTACCCA 70801 ATAGGTCGAC CAGATGCCTT GATACCCATA GTGTTCGCCA GAATTTAATG 70851 TGGATAGATC CAGTTGCGAA GAGTTAATTT TTTGTGCAGC GACATCGACA 70901 ATATAAAGAA GGGGAACTTT CTCAAGAGAG TGCGAGAGAT CCAGACTATC 70951 GTAGGGATCT TCGTTGTTGC TGTTGCGTAA GGTGAGAGTT CCTGAGATTG 71001 TGATTGTCGG GTTGGAATCT TCAGTATAGG TAGATCCTGT TTTTGTGGGG 71051 TAAATCCAAA TTTTTGGAGC CTGAGCTTGA AAAGAAAGAA TAGAAGGAAG 71101 GTCAATGGCA ATGTGATTTA AAGTTATAGT ACTTCCTACT GTCGTTGGTG 71151 TTGAGGATGG TGTGGGAATC GTTCCTGCTG TCGTGATCAC CGCACCTTGA 71201 CCTAGAAGTA GAGTGCCTCC TCGTTGGAAG AACTTATAGC AGGCCAGCCC 71251 CGCACCATCT TCAACAGCAA GGACTCCTTG ACGTAGTTCC GAAGTGTTCC 71301 TTAAATAGGA AAAGAAATTC ATTTCATCGG TAAAGTTCTG GTGTACATGT 71351 TCCCCTGAAA ATAAAACTGT ACCTGTATGA CCGGTTTCGA AATTAAAGAG 71401 TATGGGGAAG GAGGAAGGGA GCTCATGTTC TATGGGATCA TAGAACAGCA 71451 CTCGATAGCC GGGACGGGCT CCTATTTGCA GATTCATATT AGGAGTCGAG

71501 TGAATGGCGT TTCTGTATGG AGGATTGAGG GCATGCTTGG AGGCCGTATT 71551 ATTGTTAAAG ATAATATCTC CATTATCTGC AGATAAGATG AAGCTTCCGT 71601 TTCCAGAACC TGCTGATAAG TTGAGGAGAG CCCCTCCCCG AGTTGCAGTG 71651 TTATTTAAAA AGTATACAGG ACCATTTTCT TTGATAATGA TAGATTTCGC 71701 ATGAATGGCT CCACCGTTGC TCTGGCTTTG GTTATTGTTA AAGAGTACCC 71751 CTTCTCGGTT GTTCAATATC GTGCAAAAGG TGGTAGTAAG ACCTCCTCCT 71801 CCTGGATTGA AGTTCGATCC ATAGTTATTT GCGAACGCGG AATTTTCACT 71851 GATTTCTATG AGTTTTTTAT TCGAGGAGAT CGTGAGTGTT TGGGTAGAAA 71901 ATATGCCTCC CCCAGATCCT GAAGAGTTGC TTGTGATCTG TATAGCTCCA 71951 GAATTTCCCT CTATATTTAG AGAGTTCCCA CTATAAATCC CTCCTCCTAA 72001 ACTGTCCGAA TTTAGTGCCC GATTCTGCTT GATTATGATC GGGCCTTTAT 72051 TGTCTTTAAT AGATAAGTTC GTCTCAGTAT AGAGGGCACC CCCCTTATTT 72101 AAAGCGAGAT TGTCAACGAA AGTTCCCTGT CCTAGGTTAT TAGTAATAGA 72151 GCAATTTATG GCAAAGAGAG CTCCACCCAA TAGTGATCCC GCAGTGGCTG 72201 TAGGATTGTC AGAGACCAAG TTTGTAGTAA ATGCATAGTT CTGATTCTTG 72251 GAGATCGTGC AATTTTGAGC AGCATAAATT GCCCCGCCAG AATTGGGACA 72301 GACATTCTTC TCAAAGAAGA CATTCCCTAT ATTTTCAGAG ATCAGAAGAT 72351 TCTTACAGGT AAGAGCACCT CCATTCGACC GATAGTACTT ATAGTCCAAG 72401 ATGAAATCAT TGTGATTCCC GACAATTGCG AGATCTTGAT TTTGGTTATG 72451 AGTAAACCCT ACTTGAGGGG CTGCTTGATA TTCAGGACTT AATGTAATAT 72501 AGGTCTCCAA AGGAAGTTGG AGACCTTCAT TAGCCAATAC AAAAGTAAAA 72551 GGAAGCAACA TTCCGAAGCA AAAAAAGCGC ATACTCGTTC ACAAATAGAA 72601 AGAAATATTC TGAATAAATC AATGCTAGAT TTTTTGTCTA TAATTTGTTT 72701 GATTTTAATT TCCTTTTTTC ACAATAAAAT GCGAGTCCAA GAGTTTTAGT 72751 AAACGTACAT CATACTTTGG AAAATGGCGG CAAGTAGTCA GAGGATCAAG 72801 CTCTTTGCAT ACTTTGCTTA GAGCAGTTTT TGGTTCTCAG GGTCGTTTGC

72851 CTTTAGGGAG CTTGTGATGT TTGGGATCGG AGAGCCTCCT TATATCCCTA 72901 GGGGATGCCC TAGGAACTCT TCCGAAACAC CGAGGGTCTG TTAGAGATAA 72951 AAACAAAGGA CCATCGGGGA GACTTGTACT CATAAGAGCC ACTTATCTCT 73001 ATTCCTATAG ATTTTTGCTG ATTTTGATTA TTAGAAAATA ATAGATTTTT 73051 TCTGTTTTTT AAAGTAAAGT ATTTTTTAAA AGACTCATTT TTAATGAGTT 73101 ATTACTTTC TCTTTGGTAT CTGAAGGTGC AACAGCACTT TCAAGCAGCA 73151 TTTGATTTTA CTCGCTCCCT GTGTTCACGA ATTTCTAATT TTGCTTTGGG 73201 AGTGATTGCA TTGCTTCCTA TTATTGGGCA GTTGTATGTA GGGCTGGACT 73251 GGCTCCTCTC TAGGATAAAA AAGCCAGAAT TTCCTTCCGA TGTGGATCAG 73301 ATCGTGCGAG TAGAACACGT CGTGGGTCAC GACCATAGAA GTCGAGTTGA 73351 AGATATTCTA AAGAGACAAA GGCTCTCATT AGAGCCTAGA GACGAGGGGA 73401 AGGTTCACGG AGATCTGCCT TCAGCTCCTT TTTTTTGATA TCCAAAGTCT 73451 CAAGTTCCTA CAGTTGTTCT CTGAGGGGAC AGCTCTAAAT TTATTTCGTA 73501 TATTTGCTCC ACTACGCAAC CGTGTGACTA CAGAATACAG TCGTGCTAGG 73551 CAACCCGACC TACATAGAAT TGCCATCGTC TATATAGGAG TTCTCGATTC 73601 AGAAAGTTCC AAGATCCTAG AGCGGCTAAT CTCTTATATG AGTTGTATCT 73651 ATTCTGAATC GCAAATGTAT TTAAGATTCT TTATGGGCAA GAATGTAAAT 73701 CAAAGTGCTG TACTCTCAAA ATTACATGTA GAAAATCTGC ACATCCGTTG 73751 TGGGTTTTC AGCGAGGATG CTGTTCCAGA GAGTGAGCCC TTCGATCTCT 73801 CCATCTACGT GCACACAGAT CGTAGCTGTC CTCTCCCTAC GAAAAAACGG 73851 AGCAGCTCCT GGGAACTCCA AACTGTAGAA CTCCCAGAGT CAATATATCC 73901 ACAGTCGGAA TTCCTATTGA TGAGACCTCG AATGCTTTCG TAGACTCTAT 73951 GATGAAACAA GGAGTCGGGC AGGATGCTAA AGAGCTATAC ACATTTCTAT 74001 CTCGTGGGAA TGAGCATTAC CAACCGTGTC TATGGTTCAG TCTCGAAGAG 74051 GAACTCGGAT TCCTTTTCGA TGAAAAAATG CTCTGCGCCC CTCTATCTGA 74101 GGATCACTAT TGCCACTCGT ATCTTGTAGA TCTAGTGGAT CAACATTTAA 74151 AGGATTTAAT ATTATCGATG TTTTTAGATC CTCAGAATAT CTCAGCAGGA

74201 GAACTCCTCA AGGTCTCTAT AAACGTTGGA GATTCTTTTT CTCCTCTACA 74251 ACAGAAAGAT TTCCTCTCGA TGGTCTTACG TGATGAAACG GGAAAAAACG 74301 TCGTCGTGGT TTTTAAAGGA GTTCTCTCCT TACCCGCAAC CCAAGTCTGC 74351 AAATTAGTAG AGGAATTGAA CTCTAAGGAC TACTCCTACC TCAATATATT 74401 TTCTTGTCAC GGAGATAGTA GTCCTCAGCT TTTATTCCGT AAGGAATTAG 74451 AGGGAACTTC AGGGCGTTAT TTTACAGTGA TTTGCGCTTT ATATCTAGGG 74501 GATACAGACA TGCGTAGTTT ACAACTTGCT TCTGAAAGGA TCATGGTCTC 74551 TAGAGAGTTT GATCTTGTAG ATGCCTATGC TGCAAGATGC AAGCTCTTGA 74601 AAATCGATCA TACAAATTGG AGACCTGGAA CTTTCAGTCG CCACGCCGAT 74651 TTCGCAGATG CTGTAGACGT ATCAGCAGGA TTTAACTCAA GAGAATTTAA 74701 ACTGATTACG CAGGCGAATC AAGGGATCCT AGAGTCTGGA GAACTCCCGC 74751 TCCCTTCAAA AACCTTCTGG GAAGGATTCT TAGCATTCTG TGATCGAGTG 74801 ACTGTCACGA GACACTTCAT TCCAATGTTA GACGCCGCTA TAAAGCAAGC 74851 GGTATGGACT CATAAACATC CCAGCTTGAT AGATAAAGAG TGTGAAGCCC 74951 GTCACAAACT CTCACGAAAA AACATCGAAA GGCCCGTTCA TACAAAAAGA 75001 GATTATCGCA GACTGTTCTC CTCTTAAAGA GGCGCTCTTC CCAGGTTCTG 75051 ATGAAGATGT TCCCTCTACC TCTGAGGATC CTTCAGATGA TCATCCTTCG 75101 GATCTTGAAG ACTCTTAATT AGTTGCGATA GAATTCAATT TTTTATATAA 75151 AAACTATCGT GTTGTTCTTA TTAAAAGATA GTTAATTTTC TATCTTTTTT 75201 TAAATCTTTA TATAGCCTGC GTACGCTTTC ATTTTCAATG TTGGTTTGAT 75251 CCTATGGCAT GCTATATTTC TATTTGGATA TCTACAGTTA AGCAGCATTT 75301 TATTAGGGCT TTTGATTTTA CACGTCCTCT TGGTTCTCGG ATTACAAATT 75351 TTGCTTTGGG GGTCATCAAG GCTATTCCCA TTTTAGGATG CGTTGTTATA 75401 GGGGTAAGTT GGCTAGTTTC CACATGTTCT GCACGAAGGT TTGGGAAACC 75451 GGCATTTACT TCTGACGTTG CTAGTATCGT GAAAATAGAA AAAACTCGAG 75501 GTTATAATCC CCTTGCTTGG GTGGAACAGT ACTTGAGACA GCTTAGGGTT

75551 CGACTTCCTG AAGGAGATTT AGGAAAAATC CATGGGAAGG TCTCCAGAGA 75601 TTATGTTTGC GACAGGACTC CCCAAGAAAA TCTGAATATG GTTCCTCATC 75651 AATATCTGGG AGAGCTAGGT CGCGCGTTTT ATGGAATCCG CAACCGAGTA 75701 ACCAAGGCGT ATCAACGAGT CACTCCTCTG GAAGTCCCTT GTCTTACGCT 75751 CGTCGGTTTT GACATTTTAG ATCCCGAAGA TCAGGTGAAT TTCGTTCGTC 75801 TGGCTAACGG CATACAAACT CAGTACCCCC AAACTCAAAT AAAACTTTAT 75851 TTAATCTCTA TCCAAAAGAT ATGGAATCAG TGTGACGGTA CGATTTCTCA 75901 AGAAAAAGAA CAGCAACTCC GCTCTCTAGG TTTGGATGCT AAAATCAAAT 75951 GTGTGTCGGC CCCCGCTCTC CTGCTCCAGA AATATCTTCA ATCCGAGAAC 76001 TTGCCTTCCT GTGATCTTCT CATTAATTAT TACGGGAAAC AACAGTCCGT 76051 CAGAGACGTG GACTCTATAA AGAGTCTACT CAATCTTTCT TCCGAACATA 76101 TCCCTGCGAT TTCTGTAACC TATAGACCTG ACGATCCTTT TTATAGCTAC 76151 TATTTCTTTC CTGGTTCTCA AGGAGGAACG GCACCCGATC AGAGGATCCC 76201 TTGGAGTGAG CAGGAGCATC TTCAAACGTA TACCACCCTG TCTAACCCTA 76251 GATGTGATAG ATATGCTGTT CACTTGGGAA TGGAAGATTT TGCCTCTGGA 76301 GTATTTTAG ATCCTCTTAG GGTTTCGGCT CCTTTATCTG GAGAGTATTC 76351 CTGCCCTCA TACCTCTTAG ATTTAAAAAG TGAAGAGCTT CGTTGTTTCT 76401 TGTTATCCGC TTTTATAGAT CCCAACAATT CTGGTCAGGG AAATCCGCGT 76451 CCTATGTCCA TAAACTTTGG AAACTCTCCT TTGGGTCAGA GGTGGTCTGA 76501 GTTTCTATCT CGTGTTCTAC ATGATGAAAC AGAAAAGCAT GTGGCTGTAG 76551 TCTGCAATAA TCCACAACTT ATAAAAAAGA GTTTTCCCTC ACATTCTTTA 76601 TCTCTATTAG AGAACGAACT GGAAGAGTCA GGTTATTCTT ATTTGAATAT 76651 CGTTTCAGTG AGTCAGGAAC GCACGTGTGT TAAGGAACGT AGAATTTTAA 76701 GTTCTGATCC TTCGGGGAGG TCATTCACTG TAATCCTCAC TGATCTTCCT 76751 GAAGGGAGTT CGGATATCCG CAACTTGCAG CTAGCGTCAG ATAGGATCTT 76801 AGTTTCTAGT GCTCTCGATG CTGCTGATGC CTGTGCTTCT GAATGTAAGA 76851 TCTTAGAATA TGAGGATCCC GAGCAAGAGT GGGCGCAACA GTATGCGTCG

76901 TTCTATAGAA ACATCGACAG GGCAGGCGAT CTTCAACGTC AGGGGATTCC 76951 AGGAGAGCCT TTAGGGGTCT CAGCATCTAC GAGAGTAGTT TTAGAAAAGG 77001 ACATCGTATT CAATCTCAAT GCGGTAATCC AACAGGCCAT GTGGAAGTTT 77051 AAAAAACGGG ATCTTTTTGC TGTAGAAAGT CAGGCTTTAG GAGATGACAT 77101 GCGACGTGCT TTAGAAGGTT ATATCGGCAG CAGTCTCTTA GTTGAGGGGA 77151 CTATACAGCC TCAAGTCGCA TGTAATGTCA ATGTGAGTTT TGCTACGTTA 77201 GACGAGGCTG TGTGTGCAGC TTGTGACTCA GCTCAAGATG CACCTTCTGA 77251 GGAGAACAAT ACAGATGACT AAAGATCGCA ATCTTGTGAA CGAAATCGCA 77301 GATTGATGGG AACTAATTAG ACACACCTTT CTAAGGTGTT TGTTTTGATG 77351 AACCTTTTA TTAGTCCAGC AGAGCTCTTT TTTGAAGATT CTTCTTTTT 77401 CTTAGGTCAT TCTGGGTTTT TTGAAGGTAT CGAGGGTTCT TATTGTCTAG 77451 TTGTCTATAG AGGGTATCGA GGTTTTTTCT CTTAGGTATC CCACGATTCT 77501 TTTGTATAGA AAAATTTTAT GAAAGCTTGA ACTCTTTACA CTGACTTTTT 77551 ATTTTCAAA TAAAAACGTT TTTAAAAAATA TTATTATCAT AATTAGATAC 77601 TTATTTGTTT TAATGTCTTA TTTGATTAAA ATAACTTTGT TAAAATTTTT 77651 ATACATAAAT TTCTATTGTG GCTTGTCCAA GTATTTCTTC TTGGTTTACT 77701 GTCGTTCGAC AGCATTTTGT AAACGCCTTT GATTTCACCC ATCCCGTTTG 77751 TTCTCGGATT ACAAATTTTG CTTTGGGGAT CATTAAGGCA ATTCCCGTAT 77801 TAGGACACAT TGTCATGGGA ATCGAGTGGT TGATTTCCTG GATTCCCAGA 77851 CACACCGTTC GTCATGGAAT GTTTACTTCT GATGTCTCTA GTGCTATTAA 77901 AGTAGAACAA ACACGGGGTC ATAATTGTTT AGCTCCCCTA GAAGCCTATT 77951 TAAGTAGCTT GAGAGTCCCC ATTTCCCAAG AAGATCTAGG CAAAGTACAC 78001 GGGAGAACCC CAGAAGATCC CTTCGTAGAT ATCACACCCA CAGAAATTGT 78051 CCAACTTCTC CCTGATGAAG AACTCTCTAC TGTAGATGAG GCACTGCAAG 78101 GCGTTCGTAG TAGGTTAACC TATGCCTATA GGTCCGTAGA GAAACCTATG 78151 ATTCAAGATC TTGCTCTTGT GGGTTTTGGT CTCCGAGATT CTGCGGACCT 78201 CATAAATTTC GTGCGTCTTG CTAATGGCGT GCAGAATCAC TATCCCCATA

78251 CTAAAGTGAA GCTCTATTTA GCGAAGAACT TGGCAGATGT CTGGGACTGT 78301 GAAATTTCTG AAGAGGAAAA AGGGCAACTC CGAGCTCTAG GTTTAGACCC 78351 TAAAATAGAG AGTATATCCC TTACGAGTGC AGGTCTTCCT TCAGTGCCAG 78401 AAGTCGCTAC TGTCGATTTT ATGATTACCT GTTACGGGAA AGATCAGGAA 78451 GTCCAAGATC CCTAGGTGAT ACAACATCTT CTAAACTTTG CTCTAGAAGA 78501 GACCCCTTCC ATTTCCGTGC AATACCAAGA ACAAGAGAAG CTCTCTCCGT 78551 GCGATCATTC CCCAGAAATA GGTAAAAAGA AAAGATGGAA TAAGCTGGAA 78601 TCCTTCTCCA CGTATTGTTC TCTGTTTATG TCTGTTAAGG ATCATTATAA 78651 GCTGAATCTA GGAATTCAGA ATTCCCTGTC AGGGTGGCTT CTGGATCCCT 78701 ATAGGGTTTG CGCGCCTTTA TCTTCACCGT ACTCGTGTCC TTCCTATCTT 78751 TTAGATTTGC AAAACAAAGA GCTACGTCGT TCCCTTCTGT CAACGTTTCT 78801 AGACCCTAAA AATCTCACTA GCGAAACATT CCGTTCTGTC TCTATAAACT 78851 TTGGCAACTC TTCGTTTGGA CAGAGATGGT CAGAGTTTCT ATCTCGTGTT 78901 CTGCACGACG AGAAAGAAAA GCACGTAGCT GTTGTTTGTA ATGATGCAAA 78951 ACTTCTGGAA GAAGGATTGT CCCCAGAGGC ATTGTCTCTA TTAGAAGAAG 79001 ACTTAAGAGA ATCAGGGTAT TCGTATCTAA ACATTCTCTC GGTGAGCCCC 79051 GAAGGAGTCT CCAAGGTTCA GGAACGTCAG ATTCTAAGGC GAGATCTCCA 79101 AGGACGGTCC TTTACTGTCA TGATTACAGA TCTTCCTTTA GGTAGCGAAG 79151 ATATCCGTAG TTTACAATTA GCCTCGGATA GGATTTTAGT CTCCAGTTCT 79201 CTTGATGCCG CGGATGCATG TGCTTCGGGA TGTAAAGTCT TAGTCTACGA 79251 AAATCCAAAT GCATCCTGGG CTCAGGAATT GGAGAACTTC TACAAACAAG 79301 TTGAGAGAAG AAGGTAGTGT TTCTTTCAGA GAATATTTCA GAGCCTATAT 79351 GTGTGATAAA ATCGTGGCAC AGAAGAACTT CTTATTTACT TTAGACGCTG 79451 GTTGAAAGTC AGGCTTTAGG AAGAGAAATC AAAGTCAGCT TAGAGGAATA 79501 TATTCAGAGT ATGGTCGGGA TTTTGGGATC TCAGAGAACC AAGAAAAGCT 79551 TTAAGTTTTC TGTCGACTTT ACCCCTTTAG AGCAGGCTCT ACAAGAAAGA

79601 TGCTCTTCTG ATGATGACGA AGATGCAACA GCAACTTCGA CCGCTACAGG 79651 GGCAACAGCA TCTCCGACTG ACATGCACGA AGATGAGTAA CGTTTGTCTG 79701 ATACCTTAAA AGTTCCTTGC AAAGGGCTCC CTGAAAACTA AATTCCCTCA 79751 GAATCTCGAA TTCTCCTGAC TCTGAAACAA TCTTAGGTTT TCCTGAATAG 79801 AATCTGACTG AAATTTCTGC TCGAATCTAA GGGCTGTTTC TTATTTTACC 79851 CCTAGATGAG GATATTAAAT CCAAGCTAGG ACTTCAAAAG TAGTTGGTTA 79901 TTAGTTTATT AAAGAAAATA ATACTAAAAA TATTTAAAAG CTGTTTATTC 79951 AATTTAATTG ATATTTCTA TGTTGTTATT TAAAATTGTT TGTTTCTAAT 80001 TTTATTTTT TTGTTGTTAT GCCAATTCCC TATATTTCTT CTTGGATTTC 80051 TACCGTTCGA CAGCATTTTG TTAAGGCGTT TGATTTCTCT CGTCCCTTTT 80101 GTTCTAGGGT TACGAATTTT GCTTTAGGGG TCATCAAGGC CATCCCTATT 80151 GTAGGACATA TTGTCATGGG GATGGAGTGG TTAGTTTCTT CCTGTGTTGC 80201 CGGGATTATT ACTAGGTCCT CCTTTACCTC AGATGTCGTT CAGATTGTAA 80251 AGACTGAGAA GGCGTTAGGT CGAGATCATA TATCTCGAGT GGCGGAGATA 80301 TTGCAAAGAG AAAGGGGGAC CATAACTCCT GAGAATCAAG ATAAGGTGCA 80351 TGGGAAGTTT CCTGTCTGTC CTTTTGGTCG TTTAAAATCC GAGGAAACTT 80401 TAAAACTTAA GCCGGGAGAA AGAGAGGGAA CTTTAGATAC TGTATTTTCT 80501 AATACGTACG ATTTCTATTG TGGGTTCGAA ACTTAAAACT CCTCAAGATT 80551 TCTCGCAATT TGTGAGTCTC GCGAATGAAA CGCAGAGACT GCATCCTGAA 80601 GCGTTAGTTT GTCTGTATTT GACAGGCTTG AATCGCGAAT CTCAGATGTG 80651 CGATACAACT ACTGCAGAGA AGAAGCAGTA CCTACATAAC TCAGGTCTCG 80701 ACTCTAGAAT CCAGTGCAAA GACAGTAAAG AAGACGACGC TGGCTCTCCT 80751 GAAAATCCCG AACTTTGGAT TGGCTATTAT TCACGAGAGC AACAGCATAA 80801 TATAGACGGG CAGTATATTC AGCAGTGTCT AGGGAAGAGT GCAGATCCAA 80851 TTCCTTGGAT TCATGTTACT GAAGACACAA AGGATTTTTA TTACCCACCA 80901 AACTTTACTT CATACTCACA TACAAGACAA TCTACAGACC CAACATCGCC

80951 ACCAAGACTC CCTGAAAGTG AGGGGGATAA GGATTCCTTG TACGGACAAC 81001 TGAGTCGATC GTATCACCAT GAGTATATGC TTGGTTTGGG ATTAAAACCA 81051 GAGGATGCAG GACTCCTGAT GGACCCGGAT AGAATCTATG CTCCTCTATC 81101 CCAAGGGCAT TATTGTCATT CCTACCTTGC GGATATAGAA AATGAGGATC 81151 TACGAACTTT AGTCCTTTCG CCTTTCCTAG ATCCTGGCAA TCTTAGTAGC 81201 GAGGATCTTC GTCCTGTAGC ATTCAATATC GCTAGATTGC CATTAGAATT 81251 GGACTCGTTA TTTTTCCGCC TTGTTGCGGG TCAGCAAGAA GGGAGAAACA 81301 TAGTTACCCT TGCCCACGGA ACTCCTCGTC CAGAAGATCT TGATCCTGAC 81351 TCAATGAACA TTCTGACCAG AAGATTACAA ATGTCTGGAT ATAGCTATTT 81401 GAACATTTTC TCCTATAAAT CACGGAAAAT GATTGTAAAA GAACGTCAGT 81451 TCTTTGGAGA TCGTTCTGAA GGGAAGTCTT TCACATTGAT CTTATTTGAG 81501 GATCCCATTA GTGCAGCAGA TTTCCGTTGT TTGCAGCTAG CTGCAGAAGG 81551 TATGGTTGCT AAGGATCTCC CCAGCGTAGC AGATATTTGT GCCTCTGGAT 81601 GTTCCTGCAT TCAGTTTTCT GAGATGCAGA GTCCTCAGGC TATTGAATAT 81651 AGACAATGGG AGGCACGTGT CGAAGATGAA GCAGGAGAAG AAGCCAGAGA 81701 ACCAGTAATT TATTCTCAGG ATCAATTGAG CAGCATGCTC ACTACACAAC 81751 AGAATTTTGT ATTTTCTCTA GATGCTGTGG TAAAACAGGC GATCTGGAGA 81801 TTCCGTTCGA AAGGTCTTCT TACTATGGAA AGAAAGGCAC TAGGCGAGGA 81851 GTTCTTAACT GCGATATTTT CCTATTTAGG GAGTCAGGAG CGTAATGAGA 81901 ATATGGGGAA AAGAACTACC GAAGAACATG AGGTCGTTAT CAGCTTCGAA 81951 GAGCTAGATC GCATGGTGCA AGTCCTCCCA GCCGAAGTCC CTGCAGATTC 82001 AGGCAATGAT CCTACGCGTC CCGTTCCTAA TCCAGATAGT AACCCTGATT 82051 CCTCGCAAAA TGAAGGCAGT TAGAAAGTAA AAATACTAGA GAAATTTCTT 82101 ATCTCCAGAT GGAATCCGTG GTCCATGTAC CTAGGATTCC AGGAAGGGTT 82151 TGCCTAGGAA TATCTTAATT TCACAAACCC CAGGAGTGCA CAAACTCCTT 82201 AGAGGACTCT CTCTATATTT GTTCTTTCTA CACTAGTATT CCGTAGATTT 82251 CTGATTTCCA GGATAGGATT CTAAATAGTT GTATCTAAGC GCTTTTTACA

82301 ACCTTTCTCA GGTCTCCTCT TTCTAATTTT AAAAATAGCG AATTTCTTGT 82351 GGCTATAGGA GTCGCTAAGT ATCTTAGAGA TTGCTTTTTT CAAGTTTTTT 82401 TTAAATACTT CTTTCCTAAG TATTTCTTCC TAGTCGTGTT ATGGCTTCTT 82451 GTTTATCTGC CTGGTTTTCT ATAGTTCGTG AGCACTTTTA TCGAGCCTTT 82501 GATTTTCTT TGCCGTTTTG TGCTCGTATT ACGGAATTTG TATTAGGGGT 82551 CATCAAGGG ATCCCTGTTG TGGGTCACAT TATTGTTGGG ATAGAGTGGC 82601 TCGTTTCTAG GTATTTAGAG AGTTTCGTGA CCAAGCCGAC ATTTGTCTCT 82651 GATGTGGTGA GTCTTCTGAA AACAGAGAAA GTTGCTGGTC GCGATCACAT 82701 TGCTCGTGTA GTGGAGACTT TGAAGAGGCA GAGAGTCGCT GTGGCTCCTG 82751 AAGATGAGGA TAAGGTCCAT GGGAAGATTC CTGTGCATCC TTTCGGGGGA 82801 ATCCAACCTG TAGAAGTTCT CACTCTCTAT CCCGAAGTTC AAGATGCAAC 82851 GTTAGGGCTT GCCTTCTCTA AAATTCGTAA TCGTGTAAGA CAGGCGTATT 82901 TGCAAGCTCC ACGGCCAAAA CTGCAGAAGA TTTACATCAT AGGAAACGAT 82951 ATGAATCCTT TTGAAGTTGA CGACTTCTTG CATCTAGCCC GTCTCTGTAA 83001 TGAAACTCAA AGACTCTATC CTGACGCTAC GATTTCTCTA TATCTAACAG 83051 CTTCTGGTGG TCGCAATGCT ATGGACAAAA AGAATCGGAA GTTACTTAGT 83101 GATTGCGAAC TAAACCCCAA GATTGCTTGT TTGGACTTTA ATCAGGGTGA 83151 TGTAGTCAAA CAAGCAACTT GTGACTGTTG GATGGTGTAT CATGGGGAGA 83201 ATGATCAAGG TACGTTGAAT CAGATTCAGG AAGAGTTAGA AAAGTCAGGG 83251 GAGGAAACCC CTTGGATTCA TGTGGGGCAA AAGCCTCTTT CACAATCCTT 83301 GTGGGATTTC TCTCCATTTT CATCTTTGGA GATGAAGGGA GATAAAGAGA 83351 AAGCTCTAGA GTACTCTGAA TTAGAAAAAG AACAGCTATA TTCTCGATTG 83401 GTATACGTAG GAGAGCGCTC TTCGGTTCTT AGTTTGGGGT TTGGAGATAG 83451 TCGGTCAGGG ATCTTGATGG ACCCAAAACG GGTGCATGCT CCCTTATCTG 83501 AAGGGCATTA TIGTCATTCC TACCTTGCAG ACTTAGAAAA TCCCGGGTTA 83551 CAAAAACAA TTTTAGCGGC ATTTCTGAAT CCTAAGGAGT TGAGCAGTAC 83601 CATACTGCAA CCTATATCTC TAAATCTTAT CTTAAATAGC AAAACTTACT

83651 TAAGGCAGCA CTTTGGCTTT TTTGAGAGGA TGAGCAGAAG TGATCGCAAT 83701 GTGGTTGTCG TTGTATGTGA TTCTTGGTGG GGTACCGACT GGAAGGAGGA 83751 GCCAAGCTTC CAACACTTTA TTATGGAGCT AGAGTGTCGA GGGTATTCGC 83801 ACTTCAATAT TTTTGCCTTT AGATCTAATA GCATGTGTGT AGAAGAACGT 83851 AGGATCTTAA ATGAAAGTTC TCAAGAGAAA GCCTTTACCA TGATTTTCTG 83901 TGAGGATTCA GTATCTCAAG GAGATATCCG CTGTTTGCAT TTGGCGTCTG 83951 AAGGAATGCT TTGTGGTAAA GAGTGCTATG CTGTCGATGT CTATACGTCA 84001 GGATGCGCGA ACTTTATGAT GGAAGAAGTC TTAACTTTGG AGCGAGAATC 84051 TAATCTGTGG AATAGAAAGC ATGGTCTTTG GAAAAGAGAA GTTAGAAAAC 84101 AGAAACAAGA AGCTGCTTTG GATCAAGACG AGAGCGAGAT TTACGTTTGT 84151 AATCAGCTGA CGGCGCAACA GAACTTCGCT TGTTCTTGAG ATGCTGCAAT 84201 CCGCCAGTCT ATATGGAGAT CCCGTATGCC AGAACTTCTC TCTATTGAGA 84251 GACGGGCGTT AGGGGAACAA CTCTTTACTA CTGTACATCA CTACCTAACA 84301 ACGCAAAAAA AGATCCTCAG GGGAATCTAG AAACGCAGCA ATCCGCGCAA 84351 TTGTCTATAG ATTTCACAGC ATTAGATGAA GCTGTTGAAT CTCTAGGATC 84401 GACTCTTAGC AGAGCTCCTT CAGAAATATC TCCAATTCCA GAGGAGGAAG 84451 CTCACTTAGG AGCCAACAAA TAGAGACAAA GAAAATTCGA CGGTTTGAGG 84501 ATAACGATAC GCAATGTCTA AGCTTTGAAT CAGGATACTC TGCTTTACAG 84551 GCGTGTCTTT GTGCCTATGG TCTCTCTCT ATAACAGAGT CTCTCAAATC 84601 TAATTGTCAG AACCTATTTC CCCTAAGAAT CGATAACGTA TTTGTTAGGA 84651 AAACGTGCTC AACAAGAGCG TTTTATTTTC AGTGTACTTG ATGTCTAATA 84701 CAAATTGTTT ATTTTGTATT TTTGGCACAT CATTTAAATT CCTTGTACTT 84751 TTGAATCTAA ACGTAAATTT CTTATGACTC ATTGCTTACA TGGTTGGTTT 84801 TCTGTAGTTC GTCATCACTT TGTGCAGGCG TTTAATTTCT CACGTCCTTT 84851 ATATTCTCGA ATTACCCACT TCGCTTTAGG GGTGATTAAG GCCATCCCCA 84901 TTGTAGGGCA TCTTGTTATG GGAGTCGATT GGTTGATCTC TCATTGCTTC 84951 GAGAGGGGAG TCTCACACCC TGGGTTCCCT TCAGATATTG CTCCTATACT

85001 GAAAGTAGAA AAGATCGCGG GCCGAGATCA TATTTCTAGA ATCGAAAATC 85051 AGCTAAAGAG CCTTAGGAAA ACTATCGAGG TTGAAGATCT AGATAAAGTC 85101 CACGGGCAAT ATCAAGAGAA TCCTTATGCA GATATGGCCT CTAGTGAGGT 85151 TCTTAAACTC GATAAGGGAG TTCATGTTAG CGAGCTTGGC AAAGCCTTTT 85201 CTAGAGTTCG CAATCGCATC ACCAGATCCT ATAGTTATGC CCCTACTCCT 85251 CAGTTGGACT CTATAGCTAT TGTTGGTATA GATCTCGTCA GTCCTGAAGA 85301 ACAAGAGAAT TTAGTACGCT TGGCGAATGA GGTCATTCAA CTCTATCCCA 85351 AATCAAAGAC AACTCTATAT CTTCTTATCG ATTTTAATAA GGAGTGGGTA 85401 GGGGATATCT CCTCTGATAA GGAAAAACAG CTCCGTTCTC TAGGTCTACA 85451 TTCTGAAGTT CAGTGTCTTT CCGTCTTGGA ACCTCAGGGT GCCGAGGGCG 85501 AAGATACGAA ACACTTTGAC CTTATGGTCG GCTGTTATGG GAAGGATTCT 85551 TACTTAAGGG AGGGTAAAAT TTTACAGCAG GCCCTAGGGA CTTCGTTAGG 85601 TACTGTTCCC TGGGTGAATG TTATGCACAC ATTGCCATCT AGGTATAGAT 85651 CTCGGCTTTC CTTACCTATA AATACCGAAA AGGATAAGAC AGAGCTTTAT 85701 AAAGAGATTT CTCGTACACA CCATCAGTTG CATACTTTGG GAATGGGACT 85751 TGGAGCCCAG GATTCAGGAT TGCTCTTAGA CCGGCAACGA CTCCATGCTC 85801 CTTTATCTCA AGGGTCTCAC TGCCATTCCT ATCTTGCAGA TCTCACCCAT 85851 GAAGAGCTGA AAATTTTGTT ATTTTCAGCA TTTGTGGATG CTAAGAACAT 85901 AAGTAAGAAA GAGCTTCGTG AGGTATCTCT AAATTTTGCT AACGATACTT 85951 CCGTAGAGTG TGGCTGCGCT TTTTACTTTT AGTGTCCTAT GATGAGAAGG 86001 AGAAAGACGT AGTTGTCGTT TGTAATCATT CTGAACCTAA TATCCTCGGC 86051 CTGCCTCCTG AAGCAGTCTC TCAGCTTATT GAAGAGCTTA GCGATGAAGG 86101 CTATAGCTAT CTGAATGTAG TGCGTTGTGA TCTCTCCGGG GAGACTACGG 86151 TTCAACAACG TCTGCTATTG AATGCCGATG AAGGGAGATC TATGACGGTG 86201 GTGATCTCAG AGCTTCCTGA AGGGCACCCC GATATTCGGA ATTTGCAGTT 86251 GGCATCCGAA AGAATTTTTG TTTCTCGTGA AAAAGAAGCT GCTGATGCCT 86301 ATGCTTCAGG ATGTAAAGTG GTCGCTTTCG ATGATGAGCA TCTCCCTTGG

86351 GTCTCCAGTC ATATTGCCTA CGCGGAGGAG ATCAGAGAGA AACAAGAACA 86401 AACAATGCAA GGGTCTTTAA CTGAAGAGCA GTTAGGAGCA CTCCTCTGCA 86451 ACACAGTCTC CACAGAGAAA AATCTAGCCT TTGCTCTAGA CGCCGTGATA 86501 AAACAGTCTG TGTGGAGATT CCGCAATCCG GATCTTTTTG CTTATGAGAG 86551 AGAAGCTCTA GAGGCTTCAG TAACAGATGC TTTAGTATCT TACGTTTCAA 86601 ATTTAGACAT GATACCGTAC ACAAGTTCTC AGGGCATAGT CATAGAAGAT 86651 AGTAGTATCG TCCGTACCTC TCAAGAGCAT ACACTCATTG TGAACTGTGC 86701 AGCATTCGAT AAGTTAGCGA GCCAAATAGA GTTCTTATGC CCCAGTGACG 86751 TGTTGCCCAT TTCTGGTAAA GACCCTTTGA TTTCTGATGA TGAGGATGAG 86801 GAACTGAATC CTAAAGTTTC ATCTGCTGCA GACTCTAAAG ATAAAACCTA 86851 GGGAGTGAAT TCTACACGAG AATCGAGAGG AGAGCGAGTC TTTCAAGGAT 86901 TCATAATCCT TGTTAACGTA TGCATAAACA AGTGAAGCCA TTGCAACGTG 86951 AAGTAATCGC ATTGATGAAA GATGCTTTCC CTAGGGATGC AAAGCAGATC 87001 GTATTCCCTT CTTTCCAAAA CAGACTATAG ATTCAAAAGA TATTCTTTCT 87051 TTTCAAATAG ACTTGAGAGA GGGGGGGGG TTCTCATAGA GTGAGAATCT 87101 TGGCCTTCAT TGCTAAGTTC TTCGATGATG GATAGAGGAT TCTAAGACGA 87151 CCGGGGCTAC AGAAAACTCT AAGCAGAGCT TAGAGTTTTA AAATGTGGAT 87201 TTTAGTCCTG TAGACACTCG GTGGTTTGTA AATCCATTTT TCCCGTCAAA 87251 GGTATAGTTT AGAAAGGCCT GAGTGTCCTC GGTGAGATCT ACTACATCAT 87301 GGATTTGTAC AAACAAACCG TGGCGGCGTA GATTTGCTCC TGAGATCGAA 87351 GTGCTCTCTT GGTTTGAGAC GACAGTCACA ATATTGTGAG GGTTGACACG 87401 ATAGATATCA GGCTTGTAGG CAAGTTTGAT GGTCAGTGTA GAAGGAGCCT 87451 TCTTAAATGG TGTGAACCAT TGAGAAGAAC ATCCTATCGG TAGGGAAACA 87501 TTGTACCCTT TACCTCTACT AAAGCTACGT TGCAGATCTC CAGTTTCTGT 87551 GAACTTACTT TGCCAACCAC CTAGGAATTC TGCGGAAATA AAACCTGAAA 87601 GATCCCAAGC TTGAGCCAGA GGTCTTGTAA GAAGACACCA GTTTAGGAAA 87651 GGATGTTCTG CAGAAATAAG AACATAGTAA CTATTGTTAT GCCATTGCCC

87701 TTGAGATTTT GGAGCTTTGT CAGGTCTGAG GTAGGTGGTA TTTAGGTGAT 87751 TTTTGGAATA ACCATAAGCT GCTTTCCAGG AAATTAAGGC CTCGCTCTTT 87801 TGAGTCACGA TAGGGAATTG ACCAAAGAAC GAGAGTAAAT ACATTTGTTC 87851 TGAGCAACGT GAATCGTAGG GGTTGGCGTT AGTTTTTCCA TAAAGCTGCC 87901 CGAAAGAAAG TCCTAACGTA GTGTGGTCCG TGTAGTTCAT AGATAGCGCA 87951 GCTTGGTAGC CTCCATAGCG ACCTGAAAAG CCCTCATGTC CTTGTCTTGG 88001 TGTGTGTTCG ACATAGGCTC CTAAAGCTTT CGCGGTTATG GACAACCCGG 88051 GATGATCTAT CAAAAGAACA TCTTGGAGAA TATCAGAGAA TGCCTGATTT 88101 CCTAAGAAGG AAATCCATAA GCTGTTGCTG ACAATTTCTC CGTAACGCTC 88151 GGGATCTAAG ATATAGGTAG AACGCACGAG AGTGTCTGAA TTCCATACAG 88201 CATAGAGAT ATTTGCGCTA GGAGAGGGAC CTCCAGGAAA TCCTCCATCA 88251 GGAGCTGGAA TTAACAGGGG ACGGGACCAT GTGTAGGACC ACTTTCCTTG 88301 GTAGCCGTAG TGGCTTGGAG TCGCAATCTC CCCATCAGGA AATCCTGTCT 88351 TAGTAACGGT TGCTCCTTTG AAAACAGCGA TAGGAATTGC TACTGGAGTT 88401 TGTAATGACA CCATATCATA AAGATCTGTA ACGTCATGTT CATCAAGAAC 88451 CAGAGCTCCT GTTAAAGTGA CGTTTTTTGT GCCTGCATTT ACTGATGCTG 88501 AAACAAAATC TCTTTTTAGG AAGGAAAAAG GATCGAATGC TAACTTTCCA 88551 ATCGTAAAGT CTACAGCGGC AGGTGCTCCC GTGGGTGTTG CCAGCCCTAA 88601 GGTTCCTCCA GAGCCCAGGG TAAGCTGACC TGAGCCCTGA GTAGCGAAGC 88651 CAAGAACATT GACAACCGCA TTGTCAGTAA TCTTCAGTTC TCCACTAGCG 88701 ATCTTGACTG TTCCTAGAAG TATAGTTGTC GTGTTGGCAG GCAACAGGAG 88751 TTCTGTAGAG GAGAGTCCCT TACTTGTAAA GACTACAGAT CCTGAAGCGC 88801 CATTAGCGTT GATTGTAATG TCTTTATTAG ACGGACTTGT GGTTGGGAGG 88851 CTATGTGTAA TGGGATCATA AAATACAAGA CGTGAGCCTC CTTGTGCAGA 88901 TAGAGACACA ATCTCTCCCC CTGCTTCTAC AGTGATGGCA TTGCGGATTC 88951 CAGGTTTTGT ATTGAGCATG TTTCCTTGGA ACGCAATATC ACCTTCGGAA 89001 GCCAAAATCG TAAGTGTCGA TGTTTGCTTC GCAGGGTCTC CAACAGAGGG

89051 GCCTATGAAA ATAGCTCCAC CCTTCTCCGC GCGGTTTCTA GAGAATTCAA 89101 TAGGACCGCG TCCTTGAATA TTGAGCACTT TAGCACAGAT GGCTCCGCCA 89151 TTTCTCGCTG CAGTGTTGCT ATCTAGGAGC AAGGCACCCT GGTTCCCTAC 89201 GATGTTGCAG GTTTCGGCGT AGATCGCACC CGCATCATTT GGTGTACCAT 89251 TATAAGAGAA GGTGCACTTC CCCTGATTGT TTTTTAATTC GAAAGTTCCC 89301 GTAGGGATGC AGATGGCTCC GCCACTTCCT AAAGCATAGG TTCCTGATGA 89351 AGGGGTGACG CCTTTTAAGC TGTTCACACA GGAGTTGGCG GTGAAGATGA 89401 TGCATCCAGA ATTGTTTTCA AAAAGGAGAG AGCTTCCTCC ATAAATCACG 89451 CCGCCTCCTG ACGAAGTAAA GTTATGGGAG AAGCTCATGG TAGCGGTGTT 89501 ATTGATGAAT TTAACGACTG CCGTGGGAGA ACTGCTAATT GCAGAGCCGA 89551 AGTTGGCAAG GTTCCCAAAA AACTTGATCG ATTGGGATAT GTTTTCGACA 89601 GTGAGTGAGT CTGTAGGTTG GAGAGCAGAG GCTCCTGTAG TTACTGTAGT 89651 GATTGCGGGA GTTGCCGACG TAGTTACAGA TGCTGGAGAA TAGGCAACAC 89701 GGTTCGTAGT GAAGATCAAA TCTTTGATAG ATTGGAAAAC AATATCTTTT 89751 CCGTAGATGA GGCCGCCAAG TCCTGTCGAT TGGTTCCCTG TAAAGTTTAT 89801 AGAAGAAAA TTCTTGAAGG TCAGAGTCTC TGCGGCAGAA AGTAGCGCAT 89851 AGTTACTATT TGTAGGAGCT TGACAAGAGG TAAACGTTAA GGAAAGGCCC 89901 TTCCCTAATA AGGAGAGATT GCTGGAACTA TTGATAAAGG AACTTGAAAA 89951 ACTCGCCCC AAGAAGTTTG AACAAACAAA ATCTGAAGTG AGTAAGATCT 90001 CTTCACCCTG ATTCGTAATT GGAGGAACAA AGTTCCCTCC GAGTCTAGTC 90051 TCAGCAAACG CGCAACTTGC ACTACATAAA CAGGCAAGTA GACAAAAAGA 90101 TGAAGATTTG AAAGAAAGAG GCATGCCTCA ACCCTGTCGT TAAATAAGGT 90151 TTAAAATCGT AATTTGCTTC CTATATCTAG AGTATATTGA CGGGAGGTTC 90201 TGCGAATATC ACATCCACAG TGCCCGAAGA TCTCTAGAAC ATCATTTACT 90251 GAAGTATGGC TGGATGCTTG TACTAGCAAA GTACTTCTGG TTAGATTATT 90301 CCCTATAGAG GTCCACGTAG CTCCATTAAT AGGTAATGTC GTATCGCAAT 90351 CAGGATTGTG ACGATAGACA TCAGGAGCAT AGGCTACGAT TATAGTGTAA

90401 AAATCTGGTC GATTATGAGA ATTTTTACCA AAGCGGACGC CTACGGGAAC 90451 TGCAACGTTG AGTAGATGAC CGTGTCCAAA AATCCTCCCC TCGGGGGTAT 90501 TTTCTTGGAT GCCCCCATGA GTCGCGTAAG CAACTTCAGC TTTTACAAAG 90551 GGAATGATCT GCTTGAGGTT TAAGATGCGA GAAGAGAGA TGATGGGAAG 90601 GTTCCCTTCG AGTTCTCCTA ACCAGCAATT GTTACTCCAA GAGCAGCGCC 90651 CTTTAGGCAA TTTTGTATAT GAAGTCTTTA CTTTCTCATT GCTACGGCTA 90701 TAGGTAACTC GAGAAGTGCC TCCTGAGAAG AATCTCGATG ATCCAAACAG 90751 AGACTTGGTG ATGTTAGAGT ATACTGTAGC GAAATAAACG TTAGAATGAC 90801 CGTGACCTAC GAGGTAATCC TTAGATTTTG TAAACAGCTG TCCAAAACCT 90851 AGACTTAACG CAGCATCTGG AGTGATGCGT GTGTAGGTAT TGATGAGGTA 90901 GCCTCCACCC ATATGGCGGT AGCTGCGTGC ATCACCGGTA TGATTCGCAT 90951 GGAAGAAATT TGTAATTCCT GTGATGCTCA GTTGCTTCCC AGGGACATCT 91001 TCGCCATCAG CTGCTGACGC TTGACTTACA GCTCGTAAAT CTATGACGTT 91051 TGCCCATAGG CTATTAGGAA TGAGGGGAGC AAGACGCTCC GGATGAGGAA 91101 GATAACCGGT TTTTTTCCAA TTTCCTGTGA CCGTATGGGT TGTCGTGTCT 91151 ATGGTAAACT CCCAAGTTCC TTGATACCCA TAGGGAGATT GCTGATAGCC 91201 GTTTGTGCCG AGACTGAAGT CCGTAGTGGT TACAGTATTT GAAGTCGCTT 91251 TGAGTTCTAA AATCGGAACT TGCTGTAAAT CTTTATTAAA CATCCCGTGG 91301 TTGTCACAGC AATCTTGAGA GTTTTTCACA AGTCCTAAAG TTCCGGATAT 91351 AGTGAGAGCT CCATTGGTAC TCTGCACATT AACGACAGCC GCTTTAGTGC 91401 CATCCAAAGA ATCCAGATTG ATTACAAGCT TGTTTAAGGT GATAGCACCG 91451 TCAGTATTAT TAGCTCCATT TGTAGTTGCT AATGTGGTCC CTGCATCCAT 91501 GATGACGACG GACTTTTCAT CTTGCGTGAA GTTATGAACA TTTAAGGTAG 91551 CACCGTTTCT TAAAGCGAGA GTACCGCCTT CAAGTTCTAG CTTTTGGTTT 91601 AATGTAGATG TAGCATTTGC AGGGGTTGCT GCTTCGGTAG CAGTGAGGGT 91651 TTCTCCTGAA AAGACAATAG TCCCTGAATA CGCACCATCT GCACTGGCTT 91701 TGGGATTGAC GACCACAGTA GCGGCTGCGG ATGCAGCAGA TAAATCATCA

91751 GATGTAATCG GATCATAGAA GTATAGGGTA TAGCCTTGCG TAGCTCCTAG 91801 AGTGGCAAAC TTGGCATCTT TTCCGAAGTG AATACTATTG CGAGTAGGTG 91851 TCCCACTAGT GATGCTGAGG TTCTTGTTAA AGAGGATATC ACCTTGATTT 91901 GCGGATAGAG AGAGCTCCCC AGATTCGGGA ATAGCAATAG CGCCTCCTTT 91951 TCCAGCTGTA TTTCCAAGAA ATATTGTAGA TTTATTAGAA TCTATAGAGA 92001 TCTTTTTGCC ATAGAGGGCT CCTCCTTGTT CGGAGGCAAT GTTTTCTAGG 92051 AATGTAACGC TGTTTTCTCC AGATATAGTC AGGCTAACAC CTGTTGGTGG 92101 GGGGGTAGCT GGAGGAGTAC AGAAGATGGC GCCTCCATAT CCTAACAAAG 92151 GAGTGACTGC TGGTGGTGTA GGTGGAGGTG TAGGTGCAGG TAAGGAGTTT 92201 TGTGGAGATG CTGTATTGTT TTGGAAAGTC AGGTCGCTGT TATTAGAAAA 92251 TGTGACATTT CCGTTAGCAT AGATAGCGCC TCCTGAGCGC GAGCTATTAT 92301 TAACGAACAA GACTCCTGAG AGGTTCCCAG AGGTGAGCAT AGATCCTCCG 92351 GTAAGGTAAA TAGCCCCACC ATAGATCCCT GTAGCATTCG TTGAGAAAAT 92401 CACAGGAGCG CTATTGTTGA TGAGGTTGAT CGCTGCAGAT CCCGTGAGGG 92451 CCCCTCCATT AGAGATGGAT CCATTACCAT TAAAGAGAAG GCTCTTTTTC 92501 GTATTTCTA TTGTGATGCT TGTGCCTCGA ATGGCAGCTC CAAATCCTGC 92551 AGAACGGTTG TATTGGAATA GTATGGAGTC ATTGTTTGTA AAGAGCATGG 92601 GCGTTGTAGC GTAAATCGCC GATGCGTGAG GTATGACATT ACTCGCTGAG 92651 GTATCTGAAG TCAAAGATTC ACAGTTATCG AAGATCATCT GACTAAATCC 92701 TGAAAAACTC AAGGGACATA GTTCAGGATT TTGGGTGATT ACACTACTAA 92751 TCGCGGCTCC GTCAGCTGAA GAACGGATAT TTAAGAAGGA GAAAACCCCA 92801 CCTTTTCCTA AGATTTGTAG TGCTCCCGCC CTATTGCTAA AGCAACTGGA 92851 AGAGGTTCTG GATATGGCAT TATCAAGATT CGCAATGTAG AGATCCCCTG 92901 AAAAAATACA GAGTGTCCCT CTAGGATCAG AAAGTGTTGT GTAAGGAAAA 92951 ATCTTCCCAC TCGATCCATC AAAGTTCTCG GAAGGCATGA TAACTTCTAC 93001 AGTAAACGCT GTTGAAGCAA AACATGGCGC CAGTGTGGTA GAAATTAAGA 93051 ACTTACGAAT AGACGTTTTC ATTTGCACGT AGAGATGAAA CCAGATTATC

93101 CTACAAATAA GGGAAAGGCT GTAAAAAAAC AAGTACAATA AGACACAGTT 93151 TTAATCTCTT AATTTTGACA GCTTTAAGAT TACAGGATAT TTTAAAGGGC 93201 ATTTTCCCAT TTCTTACATT GCTTTCTTAG AAGAATACTT GATAGAAAAT 93251 GGCGATTCTA TTTTTGAAAA ATCTCAAGAA ATTCTCCCAA ACGAAGATGT 93301 TTTAGAGAAC CTGTAAAGTA GAAAATGGCG CTACGCCCGA GACTCGTGAA 93351 CATGTGTGCA TAGAGGGACC TATCTCTGAA TACAGATAGG TCCCAAAATC 93401 TTCTTAAAGG GGATTCCCTT AATTATAGTG TAGACTTAGA GATTATGGCG 93451 TAGAAGTGAT CTCAGGATTC CAAGTGAGGA AAACAGTTTT CTTAGTTGGT 93501 GTGATCCTTC TGTCTTTATC CAGAGAAGGG GGGTACTCTT CCCAGCTTAG 93551 GGACCAGAAA CCTCGAACGG CGTCATTCTC TGTTGTGACT TGGGTTCTCT 93601 CCAAAGTGGT CTCCCCTAGG AGAAGACTGT CAAAAGAAGG CCCTAGAAGT 93651 TCAAGAAGAG GAATCGTAAA GGCTTCCTTA AATTGAGGAA AGTCATAGAC 93701 GGACCAATAA GCCTCAAAGG CAACTTTAAG TTTTTCTAAT GAGACAAGAG 93751 CATCCTTGTC CTCGGCGCGA ATCCTTACAG GAACAAAGTT GTCGGTATCT 93851 AGTGGTTCCT GGAGACATAG TAAGCGTAGA GCTTGCTTCC TGCTTAAAAG 93901 AATGCAGTTG CAAGGTAACC CCATCCGATA GAGAGAGAGT TCCTCCTGCT 93951 AATGTGACAT CTTGTAGGAT TGTGGAAGTA AGATTTTCCG CACAAACTTC 94001 ATGATCATCC AGGCATAGTC CTGAGAAGCT AATTGTTCCT TCATAAGTTT 94051 CCTTTCCTTC AGGAGCATTG ATTACAAGAT CTGTAATTTT ATGCGACTCG 94101 CTATGGCTTA TAGGATCATA GAAATAAACT CCGGATTCTG AAACAGCACG 94151 TAGGTTCTTA AACTGTGCTC CAGATTGCAG ATGGATGGAG TTGTGTATTG 94201 TATTTCCGTC TTGTGATGCT GTATTTCCTT TGAAGATGAG ATCTCCGCTT 94251 TTCACGGATA TAGAGATCGA TCCTCCAGGA GCAATGGCAA TGGCTCCTCC 94301 ATTACTATTC ACGTCATGAT AAGCATGATT ATTTTCAAAA CACGAAGGTC 94351 CTCGAGTCGT GAGTGTGAGA TTGTGGGGTAG ATATGGCGCC GCCATAACCT 94401 TGGCTCACAT TGTCTCTAAA CACCAGGTAG CGGTTCCCGC TGAGATTTAC

94451 TGAAGGACGA CTCGCCTTAG AACCTAAAAG GTAGGGAGTA TAAATCGCAG 94501 CTCCACTCCA CGAAGAGTAG TTCCCACAGA AAGTCACTTC CTCACTATTT 94551 TCGATCATCA CGGAACCAAG ACTATAAATC GCTCCTTGTC CTTGAGGTAG 94601 TAGAGGTGCT GAGGTGAACG CTAAGTAAGA AAAATTAGAG AGAGTGAGAG 94651 TGGTGTCTCC AACGCGGTTC GAAATGGCAG CGCCAAAACC CTCGGTCATA 94701 AGGTTGTGAA AAGTGAAGTT GCAACGGTTG CCCATGAAAA AAAGATTCCC 94751 AGATCGATTT ATAAAAACCC CAGCATCTTC TTGATCATGC TTAACGTTGG 94801 AAATCCTCAC GTCATCTAGA AAGATGTAAG AAGTTCCTTC TGGATAACAG 94851 GTAATTTTAG GTTCTAAGCT TTTATTATTG ATAGCACCGT TATAACCATC 94901 ACTTTCATGA AGATATACAA CTTGTGCTGC TGCAGGGAGA GCGAGGAATA 94951 AAGCCGAGCA GGTAAGAAAA TTTCGAAGTA TGGTCATGGT TTCCTCGTTA 95001 AATCAATAAG GTTGAAGCAA CTTTAATAAA CAAGAAAAAA AGAAGTCAAT 95051 AAGAATAGAT TATTGTCTAT TAATTATTTA ACTGTTTTTA AAATAAAATT 95101 ATAACTAGAA ATTATTAAAA GAAATCTTTT TTGAAGAGGG ACAAATGTTA 95151 TTTTTTACAG TTTGCAAGGA AAGCATTCCC TATAGCAAAT ATTTCCCTAA 95201 AAGTATGAGA AAACTCCCTA GAAGAACTAG GGAGTTTTAG CAATCTAGAA 95251 TCGGAGTTTG GTACCAACAT CTACATTGTA GTTCCTTGAA GATCCACGGA 95301 GTTCCATAGC GTAATGTCCG AAGAGCTCAC AATTGGAGTT GTAGACGTAG 95351 TTGTTGCTAC CCCTCAGTAA AAATGCCTGT CTTGAAAGAT TGCCACCGCG 95401 AATTTTCCAA GAGTCTGGGC TCATCACAAG AGTCGCTGTA GATTGGGGAT 95451 TGTTACGATA GACATCGGAA ACAAAGAATC CTGAGAGATC ATAGGTGTAG 95501 GAATCTCCGA TATCCCCCTG CACGAATTTC GCACCCACAG GAATCGAGAG 95551 GTTAAGCAGC CTTCCAATAC TAAAACCACG GCCATCACTA GAGCTTTCGA 95601 AGAAGCTATT TTGTGATACA TAAACCATTT CGACTTTCAT CTGTGGAATG 95651 AAGGTCTTGA AAAGAGGATG TGGGTTGGAA AGAACAAAAG GAAGGTCTAG 95701 GCCGATACCA CCAGCTATAC ACTCGTTGCT CCAAGAACCT TCGGATTCTG 95751 GCAATGAGGT ATAGTGCGTT TCCATACGGT TGTCTGAATG GCTGAACGAA

95801 ACTTGGACAT CCAAGGCTAG GGGAATTTCC CTAGGGAATT TTTCTATAGC 95851 TGATTCAGAA AACTTTGCTC TTCCTAATCT CAAATAGTTT TGGGGTTGTA 95901 GGGTATGAGA GTGCTTGAAG AATAAAGTTC CACCGTAGGT TCTAGAGTTG 95951 TTGTGAGCGA TAAAACAATC TTTGTCTCTA GCAAAGAGAT GGCAGAACGC 96001 AAAGGTAAAT AGGTCGTCTT TAGGAGTGTG AGCACTŢCCA CCGATGACGT 96051 AGCCTCCAGA GGTATGACGG AAGCCTTTGC GATTTTCATC TCCAGTCTTA 96101 TGCAGGAAGT TCGTCATGGA GGAAACCCAG AAACCTTGTT TGTGTTCCAT 96151 ACCAGTTGCG CCGATCTCTA CAAGCTGTTG CAGAGAGCGA ATGTCAGTAA 96201 AGACTCCCCA TAGGGTATTG CATACTAACG CAGATTTTCT TTCGGGGCTG 96251 GGAACAATC CTGTTTTGGT CCAAGTTGCC GTGGCCTCTT TTGTATTTGT 96301 AGCTGTATCC GTAGTCCAAT TAACATTCCA TTGTCCTTGG AATCCGTATT 96351 CTGAATTAGG ATCCTCAGCA GGAACAGGGA TAAGGCTGCT GATGTCAACG 96401 TTAGTATCAA CATCAGCATC AACCGTGATT TTTAATAGAG AGAAGAGCTG 96451 GTCATGGCTG AACATATGAC TTTCATAAAT GTTCCCTTCA ATATCAATCA 96501 GGTTGAGCTT CCCAGATACG ATCACTTTAT TTGAAGCACC TTTTGCTGTT 96551 AGGCTGACGG GCTGCTTAAG ACCTAAGGAG TCAACATTGA TTCCTAGGTT 96601 CGTGATTGTA ATACTCCCAG CTGTAGTTGA TAATGTCGTT CCTGAATCCA 96651 TGCCGAGGAG AGAACCGGCC TCTTGAGAGA AGCTCGTGCT CTCTAAAGTG 96701 ACTCCCTTTT GTAGCAATAA CTTTCCTCCG GATAGGGAGA CTGGCTGCGT 96751 GAATGAAGAT TTTAAATTGT CAGCAACTTT AAGTTCATCT GCTGTTAGGG 96801 TTTCTCCAGA AAATAGAATC GTTCCTTGAT ATGGATTGAG AGCTCCCGCA 96851 GAGCCGTTAT TTATCTTCAA TACGTCTGAT GAGGTTCCTT CTGAAGTGAT 96901 GGGATCATAG AAGAAAATTG TATGATTTTT AGCAGCCCGT AATTCCGTGA 96951 ATTTCCCGTT ACTTCCTATG TTGATCGCAT TACGTTTAGG AGTATCGGTA 97001 CTTCCGGTTG TTGTAAGGGT ATTTCTTACA AAGGTAATGT TTCCTGTCTC 97051 TGCAGAAAGA CTGAGCTCTC CTGAGGCATC GATGCTGATA GCACCCCCCT 97101 TAGGAGTTGC TGATGAGACA TTATTTCGTA GAAACTCTGT AAAGCCTCCA

97151 GAGGAAAGGG CTAGCTTTTT AGCATGGATG GCGCCACCGC TTGTTTCTGC 97201 TACGTTTGAA GCAAAGATCA GAGTCTTATT GTTAGAGATT ATCAGTTCAG 97251 GAGATCCACT CGCCTTGGTG TTGCAGATCG CACCGCCAGT AGTTTTCGCT 97301 GCATTCCCTT CAAAATATAG AAATTTGTTG TTCGATAGTA TCGACGTGCC 97351 TTCATCATCG ATAGCGCCTC CTGACGTAGA CGCTATGTTA GATAGGAATC 97401 TAACATAACC TGTGTTATTT GCTATGCGAG CGCCTGCTGT AGTAGCAATT 97451 GCTCCTCCT TTGTTGATGA AGAGTTGTTA CTAAAAAGAG CATCTCCAGA 97501 AGTGCCAGTT AAAAGGAAAG ACGCTCCTTT GATAGCTCCA CCATCTGCAG 97551 TAGAAAATT CCCAGCAACT ACAAGTTTAC GAATATTTC TAAATTTACG 97601 CCTCCTGCTG AGGAAAGCGT TCCCTGACCT GTAGTAACCG TTGTGCTAGG 97651 AGAGGAATCA AAACTCAGTA AGGAAAACCC TGAGAAGGTA AGATTCTTAT 97701 TTGCTGTTGT AGATGCAGCA GCACCTGCAT GAGTGCCAGC ATCTATAAAG 97751 CCAAACGTTA AGCTATGACC GTTCCCCAAG AAGGTAAGAT TGTCCGTGGT 97801 TTGCTTAAAA CAACTGTCAG ATAAGGGAGT GCCTTTTCCA GGCTCGTAAA 97851 AGAAGACATC TCCTGTTAGA GAATATGTTG TGGCTGAAGT TTTTGGAGTA 97901 AACGTTCCTG AATCGATATT TCCATTAAAG CTATCATCAG GTGATAAAAG 97951 TTCCTCGTTA GCTAGTGACT GTAGGTGACA TGAGAAAGCT AACACGGAGG 98001 AAACTAAAAC CCAAGGAATC GAAGTCTTCA TGGTAATGCT TTTGTTTTTT 98051 AGAGAACTAT TCGCATCAAT ATAGAAACAA AATAAGTAAA TCAAGTTAAA 98101 GATGACAAAA CAGCTGTCAA GAATTTTTAT CTTGACTCTC TGAGTTTTCT 98151 ATTTTATATG ACGCAAGTAA GAATTTAATA ATAAAGTGGG TTTATGAAAT 98201 CGCAATTTC CTGGTTAGTG CTCTCTTCGA CATTGGCATG TTTTACTAGT 98251 TGTTCCACTG TTTTTGCTGC AACTGCTGAA AATATAGGCC CCTCTGATAG 98301 CTTTGACGGA AGTACTAACA CAGGCACCTA TACTCCTAAA AATACGACTA 98351 CTGGAATAGA CTATACTCTG ACAGGAGATA TAACTCTGCA AAACCTTGGG 98401 GATTCGGCAG CTTTAACGAA GGGTTGTTTT TCTGACACTA CGGAATCTTT 98451 AAGCTTTGCC GGTAAGGGGT ACTCACTTTC TTTTTTAAAT ATTAAGTCTA

98501 GTGCTGAAGG CGCAGCACTT TCTGTTACAA CTGATAAAAA TCTGTCGCTA 98551 ACAGGATTTT CGAGTCTTAC TTTCTTAGCG GCCCCATCAT CGGTAATCAC 98601 AACCCCCTCA GGAAAAGGTG CAGTTAAATG TGGAGGGGAT CTTACATTTG 98651 ATAACAATGG AACTATTTTA TTTAAACAAG ATTACTGTGA GGAAAATGGC 98701 GGAGCCATTT CTACCAAGAA TCTTTCTTTG AAAAACAGCA CGGGATCGAT 98751 TTCTTTTGAA GGGAATAAAT CGAGCGCAAC AGGGAAAAAA GGTGGGGCTA 98801 TTTGTGCTAC TGGTACTGTA GATATTACAA ATAATACGGC TCCTACCCTC 98851 TTCTCGAACA ATATTGCTGA AGCTGCAGGT GGAGCTATAA ATAGCACAGG 98901 AAACTGTACA ATTACAGGGA ATACGTCTCT TGTATTTTCT GAAAATAGTG 98951 TGACAGCGAC CGCAGGAAAT GGAGGAGCTC TTTCTGGAGA TGCCGATGTT 99001 ACCATATCTG GGAATCAGAG TGTAACTTTC TCAGGAAACC AAGCTGTAGC 99051 TAATGGCGGA GCCATTTATG CTAAGAAGCT TACACTGGCT TCCGGGGGGG 99101 GGGGGGTATC TCCTTTTCTA ACAATATAGT CCAAGGTACC ACTGCAGGTA 99151 ATGGTGGAGC CATTTCTATA CTGGCAGCTG GAGAGTGTAG TCTTTCAGCA 99201 GAAGCAGGG ACATTACCTT CAATGGGAAT GCCATTGTTG CAACTACACC 99251 ACAAACTACA AAAAGAAATT CTATTGACAT AGGATCTACT GCAAAGATCA 99301 CGAATTTACG TGCAATATCT GGGCATAGCA TCTTTTTCTA CGATCCGATT 99351 ACTGCTAATA CGGCTGCGGA TTCTACAGAT ACTTTAAATC TCAATAAGGC 99401 TGATGCAGGT AATAGTACAG ATTATAGTGG GTCGATTGTT TTTTCTGGTG 99451 AAAAGCTCTC TGAAGATGAA GCAAAAGTTG CAGACAACCT CACTTCTACG 99501 CTGAAGCAGC CTGTAACTCT AACTGCAGGA AATTTAGTAC TTAAACGTGG 99551 TGTCACTCTC GATACGAAAG GCTTTACTCA GACCGCGGGT TCCTCTGTTA 99601 TTATGGATGC GGGCACAACG TTAAAAGCAA GTACAGAGGA GGTCACTTTA 99651 ACAGGTCTTT CCATTCCTGT AGACTCTTTA GGCGAGGGTA AGAAAGTTGT 99701 AATTGCTGCT TCTGCAGCAA GTAAAAATGT AGCCCTTAGT GGTCCGATTC 99751 TTCTTTTGGA TAACCAAGGG AATGCTTATG AAAATCACGA CTTAGGAAAA 99801 ACTCAAGACT TTTCATTTGT GCAGCTCTCT GCTCTGGGTA CTGCAACAAC

99851 TACAGATGTT CCAGCGGTTC CTACAGTAGC AACTCCTACG CACTATGGGT 99901 ATCAAGGTAC TTGGGGAATG ACTTGGGTTG ATGATACCGC AAGCACTCCA 99951 AAGACTAAGA CAGCGACATT AGCTTGGACC AATACAGGCT ACCTTCCGAA 100001 TCCTGAGCGT CAAGGACCTT TAGTTCCTAA TAGCCTTTGG GGATCTTTTT 100051 CAGACATCCA AGCGATTCAA GGTGTCATAG AGAGAAGTGC TTTGACTCTT 100101 TGTTCAGATC GAGGCTTCTG GGCTGCGGGA GTCGCCAATT TCTTAGATAA 100151 AGATAAGAAA GGGGAAAAAC GCAAATACCG TCATAAATCT GGTGGATATG 100201 CTATCGGAGG TGCAGCGCAA ACTTGTTCTG AAAACTTAAT TAGCTTTGCC 100251 TTTTGCCAAC TCTTTGGTAG CGATAAAGAT TTCTTAGTCG CTAAAAATCA 100301 TACTGATACC TATGCAGGAG CCTTCTATAT CCAACACATT ACAGAATGTA 100351 GTGGGTTCAT AGGTTGTCTC TTAGATAAAC TTCCTGGCTC TTGGAGTCAT 100401 AAACCCCTCG TTTTAGAAGG GCAGCTCGCT TATAGCCACG TCAGTAATGA 100451 TCTGAAGACA AAGTATACTG CGTATCCTGA GGTGAAAGGT TCTTGGGGGA 100501 ATAATGCTTT TAACATGATG TTGGGAGCTT CTTCTCATTC TTATCCTGAA 100551 TACCTGCATT GTTTTGATAC CTATGCTCCA TACATCAAAC TGAATCTGAC 100601 CTATATACGT CAGGACAGCT TCTCGGAGAA AGGTACAGAA GGAAGATCTT 100651 TTGATGACAG CAACCTCTTC AATTTATCTT TGCCTATAGG GGTGAAGTTT 100701 GAGAAGTTCT CTGATTGTAA TGACTTTTCT TATGATCTGA CTTTATCCTA 100751 TGTTCCTGAT CTTATCCGCA ATGATCCCAA ATGCACTACA GCACTTGTAA 100801 TCAGCGGAGC CTCTTGGGAA ACTTATGCCA ATAACTTAGC ACGACAGGCC 100851 TTGCAAGTGC GTGCAGGCAG TCACTACGCC TTCTCTCCTA TGTTTGAAGT 100901 GCTCGGCCAG TTTGTCTTTG AAGTTCGTGG ATCCTCACGG ATTTATAATG 100951 TAGATCTTGG GGGTAAGTTC CAATTCTAGG AGCGTCTCTC ATGTCTCAGA 101001 AATTCTGAGA GAGATCGCAT TTAGGATTTT CTTAAACACG ACTCACCTTG 101051 TTTTTGAACC AGGAGAGATC GGGGATTAAA AAGGCAAGAG GGCAGAGTTC 101101 GTGAGGTCAC GTACTCTGCC TTTCTTGTTA CAAACACGTT TTAAAATTAA 101151 GGAAATTTTT TAATAGAAAC CCGTTCTTTA AAATACGTTT CTTTAATTCT

101201 TATTGAATAA GATAATTCAC TATTTTTAGA TCCTAAATTT TAAGTGGTTT 101251 TTGTTATGCT TCTTATAGAG AATAGCTGCA AAGATTAGAG TTGCAGAGAC 101301 GGTACGTCTC TTTCTTTTT AAGGGAAGGG GTGTTGTTAC ACCCATCCTA 101351 AGATTTGTGA GATTCCCCTC AGGCAGTAAC TTTTACAATC GTACTTTATG 101401 TTTTGATCTA GCTGTTTTCT TGTCTTTAAT TTATTCAACC ATCGAGAAGA 101451 GAGATCCATG AGTGGAAATG TATTTTATTA GGATCATCTC TAAGGATGGA 101501 AATGATGAGC CCATTCCAAC AACCTGAGCA ATGTCATTTT GATGTTGTGG 101551 GAAGTTTCTT ACGTCCTGAA AGTCTTACAC GAGCACGCTC TGATTTTGAA 101601 GAAGGAAGAA TTGTCTATGA GCAGATGCGA GTTGTCGAAG ATGCTGCTAT 101651 TCGTAATCTC ATAAAAAAGC AAACAGAAGC AGGTCTTATC TTTTTTACTG 101701 ATGGGGAATT CCGTAGGTAT AGTTGGGATT TCGACTTTAT GTGGGGATTC 101751 CATGGCGTGG ATCGTCGCAG GGACTCTAAT GACCCTGAAA TTGGAGTGTA 101801 TCTTAAAGAT AAAATCTCCG TATCAAAACA TCCGTTTATA GAACATTTCG 101851 AGTTTGTCAA AACTTTTGAG AAGGGAAATG CAAAAGCAAA ACAAACGATT 101901 CCTTCTCCAT CACAATTTTT CCATGAGATG ATTTTTGCTC CTAATCTGAA 101951 AAATACTCGG AAGTTTTATC CTACGAATCA AGAGCTAATT GATGATATTG 102001 TCTTTTATTA TCGCCAAGTC ATCCAAGATC TTTATGCTGC AGGTTGTCGT 102051 AATTTGCAGT TGGACGATTG TGCTTGGTGT CGCCTCTTGG ATATACGAGC 102101 GCCTTCTTGG TATGGTGTTG ATTCTCATGA CAGGTTGCAG GAAATTTTAG 102151 AACAGTTTTT ATGGATCCAT AATTTAGTGA TGAAGGATAG ACCCGAGGAT 102201 CTTTTTGTAA GTCTGCATGT CTGTCGTGGT GATTATCAGG CCGAGTTTTT 102251 CTCTAGACGA GCTTATGATT CTATAGAGGA GCCTTTATTT GCTAAGACCG 102301 ATGTGGATAG TTATCACTAT TATTGGGCTC TTGATGATAA GTATTCAGGA 102351 GGTGCTGAGC CTTTAGCTTA CGTCTCTGGA GAGAAACACG TCTGCTTGGG 102401 ATTGATCTCC AGCAACCATT CTTGTATTGA AGATCGAGAT GCTGTGGTTT 102451 CTCGTATTTA TGAAGCTGCG AGCTACATTC CCTTAGAGAG ACTTTCTTTG 102501 AGCCCGCAAT GTGGGTTTGC TTCTTGTGAG GGAGACCATA GAATGACTGA

102551 AGAAGAACAG TGGAAGAAGA TCGCCTTTGT GAAAGAGATT GCTAAAGAGA 102601 TCTGGGGATA AAGAATCCGG AGTTTTTATC GACTCTAAGA GTTTTCGGAT 102651 CATAGAAAAC ATTTAAATAT TCAAGAGTCT TTGGCTATTG GATCATAGAC 102701 AGTCTTAGTA TACTAAAAAG TCTTTGGATT CTAAGACGGG CAGAGTTCGT 102751 GAGATCACGT ACTCTGCCCA TTCTTTCTTG TGATCTAGCG ACTTCTTTGA 102801 ATCTTCGACC TCTTGTAATC TGGGATTTTT TCTAGTTCTT AGATTCCTCT 102851 GATCTTTCGA CTTCTCCTCG TCTAAACAAG GCGCATTGTC TTTGAGAAGT 102901 CCCTAGATAC ACTCAGGATC TCTTAGAATT TCTAAGGGAT CAGGAACGCT 102951 TTTAGAACTG GAACTTACCT CCAAGATCTG CATTGTAGCT GCGTGAAGAT 103001 CCACGAATTT CCATAGATAG GTTACTTGTG ACCTCAAGAT TTGGAGAGAA 103051 GGCATAAAAG ATCCCTGCTC TTCCGATACC AGCTTGTCTT GAGAGATTCG 103101 TTCCTGTAGT TTTCCACGAG GTATTGTTGA TTAGGAGAGC TGTCGTGCAG 103151 TCAGGATTCT TACGATAGAC ATCGGCAACG TAGATGACAG TAGCTTCGTA 103201 AGACGCACGC TCGTTTCTCG AGAATCTCTC GAAGGTAATT CCAATAGGCA 103251 CAGAGACGTT AATTAAATCA CCGCTATCGA AAGATCGTAC CAAGGTAGTA 103301 TTACGTTCTT TGAAGCTATC TTGGTGTATG TACGAAGCTT CTACTTTGAT 103351 GAAAGGAAAA TACGCGTGGA AGAGACCCTC ATGGCTTAAA GCAGTGTGTG 103401 GTAGGGAGCT CGCAAGTTCC AGAGCGCAAC CGTCATTATA CCACGAGCTC 103451 TCTCCCTTTG GTGCTTGGGT GTAATAGGTT TTCATAGTAT TTTTACTATA 103501 GATATAGCTG ATCTGAGCAT CAAAGAGGAC AGGCTGCTCA CTTTCAGATC 103551 CAGGAAGGTA GCGTAACAAG CTTGGAGAAG ACAAGGTCGC TAGATGCTGG 103601 AGATGGAGAG AAGCTGCATA GGCAGAAGCT CTATTTTTAT TTATAAAGTG 103651 ATCTCTATCT TTCCCGAATA ATTGGCAGAA GGCTGCAGTG ATAAGATTAT 103701 CAGAAGCTAA TGTTGTAGTC GCTCCTACAA CATAACCTGC ACTTATGTGG 103751 CGAAAACCTT TATTTATCTT CGTGCTATCT TTATGGAAGA AGTTCGAGAT 103801 CCCTTCACAC CAGATGCCGC GAGTTTCTTG AGATTGGCGT ACTTTAGTGG 103851 CTACAAGCTG TTGTATGGAG CGCACATCAA CAAAGGATCC CCATAGCGTG

103901 TTAGCAACTA AGGTTCCACG ACGCTCAGGA TTCGGATTGT ATCCTGTTTT 103951 TGTCCAGGTA AGAGTCGCTG CTTTGGATTT AGTCGCAGTA TCCTCTTGCC 104001 AAGATAATGC CCAATTCCCT TGGTATCCCC AATGGATAGG ATTTTTTTCT 104051 AGGGGATCAG CAGCTAAGTC TGTGATGTGA ATATTCGCGG GGTCGTCAGC 104101 AGTAAGAGTG AGACAAGAAA AGACTTGAGG GTTATTCCAA GAGACATCTT 104151 CGTAGACATT TCCAGAAGGA TCTACAAGAG AGAGCGATCC AGATAAAGTG 104201 ACTGTCTGAC TTGCTTGTGT TGCTTTTAGC GTAGCCTTCT TGGTCTCTTT 104251 TAAGGAATCT ACATTGAGAA CAAGATTATT GATAGTGATC CCATCAGCGG 104301 TTTCTAATGT GGTCCCTGCA TCCATGAGGA GGGTAGAGCC CGGAGATTGC 104351 GAAAAGGACT TAGCAACTAG AGTGACTCCT GATTTAAGAG AGAGTTGCCC 104401 TCCCGCAAGA GTTAGAGGTT GCTGAATTGT AGATTTGAGA TTATCAGCTT 104451 CTGCAGCTTC TGCTTCCGAG AGCTTCTCTC CAGAAAATAC GATGGTTCCT 104501 TGATATGCAG GATTCCCTGC AAGGTCAGGA CCATTTAAGT TTAGAGCATC 104551 TGAGAGAGCT GCAGTGATGC TAGTTGTTAT AGGATCATAG AAGTAGATAG 104601 TATTGCCTTG AGAGGCTCGC AGCTGTACAA TCTTAGCATT GGTGTTTCCG 104651 ATGTTAATAG AATTTCTGGT AGTGGTCTGA CTCGAAGAAG CTCCTTTGAC 104701 TACTGTGTTT CCTTCAAAAG TGATGTCTCC ACCAAGAGCC GAAAGACTCA 104751 AAGATCCAGA GTCAGCAATC GCAATTGCTC CTCCTAAGGG AGCTGCAGTA 104801 TCTATAGCAG AGTTGTTTTT AAAAAGCGTA GGTCCTCCAG AAGAAAGAAC 104851 TAGATTGTCA GTATAAATCG CCCCACCACT AGTAATTGCT GTATTTCCTA 104901 TAAAGTTCAG TTCCCCGTTG TCTGATAGAG TTAAGACTGG TTTGGGGGGCT 104951 GATGTACTAC TACAGTAAAT GGCTCCCCT GTAGCTGAGG TTGCGGTCAC 105001 ACTATTGTTT ATAAAGCTAA TTGCTTTGTT GCTGCTAATA AAACTGCTAG 105051 CTTCCGTGTA AATGGCTCCG CCATTGTTCG CCGCGGTATT TTCAGAAAAT 105101 GATGCTGAGT TTAACGTATT GTTAATTGTA ATCCCTCCCG TGGAATAGAG 105151 GGCACCCCT TTTTGCGTTG CTTTGTTTTT GGCAAACGTT AGGTTGGGGT 105201 TTAGCGATAG ACTGATAGAG CTGCCTTGGA GGGCGCCTCC ATTGTCATTA

105251 GAAAAGTTTT GGCCAAAGTA GCAACTATAG TTCGACTGAA TAGAACAAGC 105301 TCCTGTGGAC TTGATGGCTC CTGTTCCTGT GGTAGCATTC GTGGTTTGTA 105351 TTAGTGACAA ATAGGAGAAT CCTGAAAAGG AGAGAAGCTT ATTTGCAGCT 105401 GTATTGGTAA AGGTACAGTT CGCTCCCGCA TCGATATTTT GTAGGAGAAA 105451 TTGGTAGCCG TGGCCTTGGA AAGAAAGATT CCCAGTAGTT TCTTTAAAGC 105501 AGGAAGCGGT TAGAGCTGTC GGAGATCCTG CATTGGTGAT TGAGACATCC 105551 CCTGTTAGAT TATAGATAGT TCCATCTGCA TTTGTTGTTT GGGCTGGAGG 105601 AGTGTAGGTT CCTGGTCCAG AGAAGCTATT GGTAGGTCCT AGATTGATTT 105651 CAACAACAGC AGCAAACGCA GAGAAATTTA GTGACAAGGG AAGTGCTAAA 105701 GATGACGAGA TTAAAAAACCA ATGAAGAGAG GATTTCATGT AGAGGGCTAT 105751 AGGTGGTTTA ACAAATTATT TCACCACATA CTGCAATAAA TTAAAGAAAG 105801 CAAGAGGAAA GGAGAGACTA GTAAGTTAAG AATCTACAGG GTTTTTATAA 105851 GAATTCCTCC CTAAAAGTTT AGGGAGGAAA GTAGGAACTA GAATGAGTAT 105901 CTTAGCCCAC AATCTACATT GTAGATGTGT GCTGAGCCAC GAAGCTCATA 105951 AGCAGCTTCC CCAGAGAGTT CTACATGAGG GGAGAGAGTC AGATGGCTTC 106001 CAGCACTTGC TAAGAAGGCT TGTCGTGCGA GGTTTTTACA TAGCGAAGTC 106051 CAAGAGGCTC CACTGACCAT TAGAGAAGTA CGCGAACGGG GATTTTTACG 106101 ATACACATCA CCAATGTAGG CTAGAGAAAT CTCGAAATTA TTTTTTTCAT 106151 CTTCGGAGAT TTTTTCTAAC CGAATGCCGA CAGGGATAGA GCAGTTCACT 106201 AGGTCTCCAT CATCAAAAGC ACGGGCTTCA GCGCCACTCT CTTTAAAGTT 106251 TTGTTGGCGG CTGTAGACTG CCTGGAACTT TAAGAAGGGG AAATATCCCT 106301 GGAAGAACGG TGCTTCTTTA GGGAGATATA GAGCCAGAGA TCCTCCGAGC 106351 TCTAGAGCCC CAGAGTTATT GGTCCAAGAG CCTTGAGCTT CAGGATAGGA 106401 AGTATAGCGA GTATCCATAT CATTTTTAGT GTAGCTGTAG CTTAGCTGGG 106451 CATTCAAAAT GAGAGGAATA TCTTTCAGCA TGTCGGTGAT ACTTCCAAAT 106501 GAGGGCATGG GAAGTCCTCC TAGGAATGCT CGATGTTGCA GGTATAGCGA 106551 CGCTAAATAG TTATGAGAGG TATTTTCAAC TATAAACAGG TCTTTATCTT

106601 TACCGAAGAG CTGGCAGAAA GCTACACTGA AGATATTTTC AGAAAAATCT 106651 TCAGCACTTC CTCCAACAAT ATAGCCGTAG CTTTTATGTC GGAATGCTTG 106701 GTTAGTTCCT GATTTATCCT TATGGAAGAA ATTCGCAGTT CCTGATGCCC 106751 AGAGTCCTCG TTGCTGATAG ATACTATTCG CTTGAGATGT CATGATCTGC 106801 TGTAGAGTGC GAATGTCAGT AAAGGATGCC CATAATGAAT CGGGAACTAC 106851 GGAAGCTCTA CGCTCAGGAT TAGGGTTGTA GCCCGTAGTT ACCCAAGTCA 106901 TAGTTCCTGA TTTTGCAGTT GATGTGTCTG CCCAAGTGGC TTCCCAATGT 106951 CCCTGATACC CGTAATGAGG TTCTGGAGTT TGTACTGGAG AAGTGAGAAG 107001 CGCATCGATA TAAATATCGC TAGCAGCAGT AGCAGCAGTG AATACCACCA 107051 AAGGCTGCGT GAAGGCTTGG TTTATCGTAT GGCTTTCATA AAAATTGCCG 107101 CTACTATCTT GGAAAACAAG AGGAGAGGTT AGAGTTATAG TTTTGTTGGC 107151 TCCTGCTGTT TCAATGGACA CACTCTTATT TCCCTCTAAG GCAGAAAGAT 107201 CAACGACAAG TTTGGTAAGA CTGATAGCTT CAGTATCTGC TTTGAGCTTT 107251 GTTCCTGGTT GCATGAGGAG TGTAGAGCCT TCAGTCTGTG TGAAACCATT 107301 GACATCTAAC TCGACATTTC CTTTGAGTGC TAAGGTTCCA GAGGCTAGAG 107351 CCAATGGTTG CTTTAATATA GATGTGAAGT TATCAGCAGC TTTCGCTTCA 107401 TCTGCAGAGA GCTTTTCCCC AGAAAATACA ATCGTTCCTG AATAATCTAA 107451 AGGCGAGTTG CTATCCGGTT GGTTGATGGT CAGAACGTCT GAAGCTCCTG 107501 TGGTGTTAGA TGCAATCGGA TCATAGAAAT AGATAGATTG GCCTTGGGCT 107551 GCCCTTAAGT TCGTAATTTT TGCTGACGAT CCCAGGTAGA TAGCATTCCG 107601 TGTCGATGTT GGCGCGGAGG TTGAGGTTAG AGTGTTGCCA AGGAACGTGA 107651 TGTCTCCTTG ATTTGCAGAG AGACTTAAAG ATCCAGAGTC GGCAATTGCA 107701 ATAGCGCCGC CCTTGCCTGC AGCTGTGTTC CCGCATCTAT TATTTGAAAA 107751 TAGGGTAGGG CCAGCAGCGG AAAGATCTAG ACCATGGGCA CAGATTGCTC 107801 CGCCTTGAGT TACTGAAGAG TTCTCGGCGA AGGTCAGACT TTTATTTCCA 107851 GAGATAGTAA GAGTAGGAGT CTCTCCTGTT TTTTCACAAT AAATGGCCCC 107901 GCCCTTGCCT GCAGCATCTG TTGCAGTGTT TCCAGAGAAG AAAAGGGAGC

107951 TATTTTGAGT AATCGAGGAG CTGGCTTCAA AGCCCAGAGC CCCACCCCCA 108001 GTTTCTCCTT TATTATTCAT AAAGACTAAC TGGCCGGTGT TTCCTGAAAT 108051 ACTTGCAGCC GCAGAGCTAT AGATCGCTCC ACCTAATTTT TTTGCGCTAT 108101 TACTAGTGAA GGTTATAGAA GAGGTATTCC CAGAAATAGA AAGAGTTTTT 108151 GTGGTGATCG CTCCGCCATT GTTATTAGCT TCATTGGAGA CGTTTTGGCT 108201 AAAGAGAATC GTTCCATTAT CGGTAAGATT TAAGGCTCCT GCAGAACTTA 108251 AAGTACTTTT TCCTGAAGCA ACTGTAGTTC CAGGAGCTGC AATGAAGGAA 108301 AGGTTAGAAA ATCCTGTGAA TGTTAGGGCT TTATCAGCAG TTGTGCTTGC 108351 CGCAGCTCCT GCATTCGAAC CCGCATCTAC CGTGTTGAAT GAAAATGAGT 108401 ATCCCTTTCC AGTAAATGTC AGATCACCCG TAGTTTCTGT AAAGCAGCAG 108451 CCTGTTAATG CTGTGCCTTT CCCAGCATCG TTTATATAGA CATTTCCTGA 108501 TAAGACATAG TTCGTTCCAT TGGCATCTGC TGTAGATTTT GGAGTAAATG 108551 TAGAGCCGCC CGCTCCATCA AAGCTATCTG TAGGGGATAA AGAAGCATCT 108601 GCTCCGTAAG TTGCAATGCT CAATAGAATG GGAGTGACAA GAGTCGAAGA 108651 GATCAGGAGT TTGTGCAAGG GTATTTTCAT AGAAAGATGC TTGGGTTCAA 108701 TTAATTAACA CGTTTTCGAT AATCTAGAAA CAAAACTTAG AGCCTAGGTT 108751 TGTATTATAA TTTCGTGAAG AACTTCGTAC TTCAAAAGCG AATTGACCGA 108801 AGATTTCCAT GTGGGGGTTC ACTTGGAAAT GGTTCGCAGC ACGAACAGAA 108851 AAACCTTGTC GTGCGAGGTT GGTACCATAG GCCATCCAGT TAGCATCGCT 108901 AGCTATTAGG GAAGTTTGAC ATTTAGGATT GCGTCGGTAA GCATCGAGTA 108951 TATACATAAG AGTAAGATCG TAAGTTCCCT TTTCTGATTT TGAGTCTCTT 109001 TCGAAGGTGA CGCCTATAGG AATCTCTACG TTGATAAGCT CGCTTTTATT 109051 GAAAGCGCGT .CCTTCAGCAT GACGCTCGTA GAAGTCTTGC TGATGCGCAT 109101 AGATATACTG TACTTTGACA AAAGGTTCGA CTTCTTTCAG AAGATACGGA 109151 ACGGAAATAA CAAAAGGCAG GCTAGCTCCA AGATCTGCAC AGAAGGCATC 109201 GTTTCTCCAA GAACCCTTGA TGATAGAGTT ATCGGTATAA TATGTCTTCA 109251 TGTGGTTGTC TGTATGGAGA TAACTGAATT TAGCATCGAA CGATAAAGGA

109301 ATGATCTGGG AGATCTCAGA GAGCACCCAG GGAGCTCGGG TTGCTTTTCC 109351 CCAGAGGAAA TTGGCGATGT CGAAGAGCCC TTCTGTATGG TGGAAATACA 109401 AAGAGGCACC GTAAGTATCT CCGTGGTTCT TACCTGTAAT ATGATTGCGA 109451 TCTCTAGCAA AGAGCTGGCA GAAGGCAAAA GTAAGCTGAT CCTCGGCAGG 109501 AGTTGTTGCT GTGATCCCTA GTGCATAACC CCCGCTGATA TGGCGGAAAC 109551 CATGGCGGGT GGGCATAGAA TCTCTATAGA AGAAATTCGC AATTCCTGAA 109601 AGCCATAGCT CACGCTCAAA AGGCTCCCCA CTGGACTTGG TTTCTATAAG 109651 CTGATTGATC GAGCGTATAT CTATAAAGTT TCCCCATAAG CTATTTAGAG 109701 GGAGATTACT TTTTCTCTCA GGACTAGGAA TGTATCCTGT ACGGGTCCAG 109751 TTGATGCTTC CTATTTTGA GGATGTTGCA TTTGCCCAAG ACAACTGCCA 109801 GTTTCCTTGA TACCCGTAGT GGGTTTCAGG TTCTTGAAGA GTCAGGGTAG 109851 AAAGAGCTCC CAGAGTAATC GTTCCGTTGG CTCCTGCGGT GGTAAGTTCA 109901 AGAAGAGGAT AGGTACTAGC ACTTTTTAAG TTATGATTCT CATAGAATGA 109951 CCCTTCCGTG TCAATAAGCG CAATCGTTCC CGATAGGCTG ATATTTTTAT 110001 CTGCAGCTTC TGTTTTTAAA GCTGCCTTGT TGGTTCCATC TAAAGAGGAG 110051 AGATTTACTG CTAAGCCATT AAGCGAAAGA TTTGCCTCTT TAGCACTAAG 110101 TGTAGTCCCC CCATCCATTA AGATGCGGGA TCCTGGACTT TGAGTCAGAT 110151 CCTTGAAAGT TACGGTGACT CCATCACGAA GTACAAGATC TCCCCGCGCT 110201 AATACTGCAG GTTGTCGGAT AGTAGAGGTG ACGTTTGCAG CGATTGCTTT 110251 TTCTGTAGGG GAAAGCTTTT CTCCAGAAAA GACAATCGCA CCCCCATACT 110301 CGATCTCACT GTTCGCATCT GCTAAGTTTA AGTTCAATGT GTCGGTAGAA 110351 GCTGCGGTTC CTGGATTTGT GATGGGATCA TAGAAATAGA TAGATTGCCC 110401 CGTAGCAGCT CGTATCGATG TGACTTTAGC GGTATCAATG ATATTTATTG 110451 CGTTTCTTGT ACTTGTGCTT CCGTTGGTGA CTTGGTTGTT ATTGAAGGTA 110501 ATATCTCCAG AAGTAGCAGA GAGAGCGAGT TCCCCAGCAG ATGCTATATT 110551 GATCGCTCCT CCTCCTCCT GACCGGCGCT ACTTCCTGAG ATATTACTTT 110601 GAAATAGAGT AGGACCTCCA GCGGAAATAC TGACCTTGAG TCCAGAGATG

110651 GCTCCGCCAT ATGTCAATGC TGTATTATTT GTGAAAGAGA GGTTTTTGTT 110701 CCCAGTAAGA GTCACTGTTT TATCTGTCGT AGTGCAACAA ATAGCCCCGC 110751 CCTGAGCTTG AGCGGCTTCC CAAGCACTAT TGCCGTCAAA GATCACTTGA 110801 AAGTTATCTG TAATCGAACA GTTGTCAGTG CTGTACAGAG CACCGCCAGA 110851 TCCTTTCGCT AGGTTTTGAG AGAAGGAAAC TATCCCAGGG CTGTTCTCGA 110901 TAGTTATAGT TCCTGTAGCG TAAACTACAC CGCCTTGCTT CCCTGTGAAG 110951 GCTTGGTTTC TCGAAAAGCT CGCAAACTGA GATGTCCCTG ATAATAAGAA 111001 GTTTTCGTA TTGATAACAC CGCCGTTATC TGACGAGAAG TTCTGAGTAA 111051 ATATAATTTG GGAATTGCCA GTTAGAGATA GATTCCCCAC AGATTTTAAA 111101 GCACATTGTC CAGTAGGAGA GAGAAGAAGA GAGGGACAAG AGATAATAGA 111151 GAGTCTAGAA AAATCATTAA AGAGAAGATT CTTATCTGCT GCTGAGGTAC 111201 TGGCTACAGT TCCAGCGCTA GAGCCCGCAT TGATAAATGC AAACTTCAGT 111251 GCATGTTGAT TTCCTTGGAA AGTAAGATCG CCGCCCGCTT CTAGGAAGCA 111301 TCCTGAGGCT AAGGGAATTC CTAAAGCCCC TGCATTTTGA AAGGATACGT 111351 CGGAAAGTAA GGAATAGGTA GTTCCTGCAG CAGCGTCCGT AGTGGAAAAG 111401 ACCGTGAAGG TAGTTCCGTT AGATCCATCA TAGCTATTAT TGCTGCTATC 111451 TAAGGTCACC TCTGCCGCGA CTATAGAGAG CGATGAAAAG AGCGGGATTG 111501 AAGAAAAGAA CAACCAAGAG ACAGAGGACT TCATTTGTAA GCACTTTTTT 111551 GAAACAAGGA AATTAAATTA GCAAATACTG TAAAGAAAAA AAGAAATCAA 111601 GGGAAACGCA AGGAATTGAT TGATGCGGAG AATCAGAACC CCAAGGATGG 111651 CGGATCTTTT ACTTCTCTTC ATACGGATCC TAAGAATCTC TTTGATGAAG 111701 AGGGGATGCC CTCCCCCTCT GATACCCTAC AGTGCGATCT CAATAACGTA 111751 TTCATCTTTA TAAAAAGTAT GTTTTTCTAA GATTCTCGGA GAATCTTAGA 111801 AAGAATAACG AGTTCCACAG TTTGCATTAT AGCTTCTTGA GGAGCTGCGC 111851 AGTTCACAAC TTCCAGAAGC GAAGCAGTCA AGACCATGAA GTAACTTCAG 111901 ATGTCCAGAA GCCTCAGCAA AGAAAGCTTG TCGTGATAAG TTTGTAGCAA 111951 ACGTAGACCA CGAGGTGCCA TTTGTTAAGG AGGTCAGGCA GTGAGGGTGA

112001 TCCCGGTAAG CATCTACAGC GTAACCTAAA GTAAGAAGCA AAGCACTGGG 112051 GGGCTTTGCT GATTCGTGTT TGAAGGTGAG TCCCATAGGG ATAGACACGT 112101 TGACCAGATG GCTAGCGTCA AAGATACGTG GATCAGCAGC AACCTCTTGG 112151 AATCCTTTTT GATTTACACT CACAACTTGG AGTTTCACAT AGGGAGAGTA 112201 GCTGGTAAGG TATCTGTAGT TTAGATCTAC AGGAAGAGAA CCACCGACTT 112251 CAACAGCGAA GCTATGGCTG TCCCAGTCTG ATTTCCCTTG TGTGTTGTTC 112301 GCAAGCTTTG TCGTCATATT ATGGTGGTTT CTTCCATAGG AAACTTGACC 112351 ATGGAGAACA AGGGGAGTTT CTCCTGGGAG CTCTGGAAGG ACCTTAGAGA 112401 GGACGTGGCG ACGTAATGAG CTATGCAGGG GAATGACATA AGAGCTCTGA 112451 GCACAGAGAG ATCCTGCATA GACTTGAGAT TTAATATCCG AGACTACGTA 112501 ATCCTTAGAT TTGCCAAAGA GTTGGCTGAA TGCAACAGCA AAGGTATATT 112551 CTTGAGGGGT GGTCATGCTG CCACCAACAA TATAACCTCT GGAAATCAAA 112601 CGGAATCCTG CATTTTCCTT TTGCTTGTCT TGATGGAAGG CGTTGCCAAT 112651 ACCTCCAATC CAAATCCCTG GATGTGAGGG AGCGTCCGAC ATCGCAGTGG 112701 CGATCTCCTG CTGTATAGAA TGGATGTTTA CATAAGCATT CCAAAGGCTA 112751 TTAGGAACTA AAGTCGCACG AAGCTCTGGT TTAGGAGTGT ATCCTAACGC 112801 TTGCCATTCC GCGACCAAAG TCACCTTCCC TCCAGCTCCT ACTTTAGGAA 112851 CCAGAGTCCA ACTCCCTTGA TACCCATAAT CCGGAGCAGC CATGCTAGAA 112901 GGAATCGGAT TGAAGTCGTC TAAATTTACA GTTCCTGAAG TAGAAGAAAG 112951 ATCTAAGAAA GGAAGATTTA AGTTTGCTTT CAACCCAGGA TTGTCATAGA 113001 AACTTCCTTC ATTGTTATGG AATTTCAGAT CCCCTGAGAT TTTTAATCCC 113051 CCACTTGTGC TGTTTACGGC AATCGTTATC ATACGCTTGC CATCTAAAGC 113101 ATCCAGATTT ACAGAGAGAT TCTTTAGATC GATGCTGCCA TCTGTATTGT 113151 TAGTTGTCGT GGTCTCTAAG GTCGTTCCTG CATCCATGAA TACTGTAGAA 113201 TCAGGCTGCT GTGTGAAGGA ATATACTTGT AGGGTGGCTC CTTCTTTTAA 113251 AACGACATTT CCTCCTGCTA AGTTGATCTT CTGGTTCAGT ATGGTGGTAG 113301 TATTTGCAGG AATCGAGGCA TCTTGACTGG GGAGTTTTCC AGAAGAAAAT

113351 ACTATAGTTC CCGTGTTTGG GTTTGCAGGT GCTACAGGGA CTACAGGCAC 113401 TGAAGCTATA GGACCATTTT TTGGTTGGGG AGGAGGAACA ATAGCTTTGA 113451 CAACAGGATT GATGACTAAC TCCTCTATTG TTCCTCCAGA TGCAGGAGCT 113501 TCCATCGTAA TAGGATCATA AAAATAAATC GTATGACCAG GAGCTGCTGC 113551 AAGCTTAGTG ATCTTAGCCC CTGCACCTAA ATGGATCGAG TTGGGAGTTG 113601 AAGTTCCCTC AGTCGCTCGG TTCCCTGAGA AAGTAATATC CCCATCAATA 113651 GCCTCTAAGG AAAGTTCTCC GCTATCGGCT ATATAAATGG CGCCTCCCTT 113701 GCCTCCAGAA TTATTGGTAA AGGAGACAGG ACCGTTAGCT GTAATCGAAA 113751 GGTTTTTCGA ATAAATCGCT CCTCCCGAAG TTTCAGCAGT ATTGCCATCA 113801 AAGTTTATGG ATTCACTGCC TGAGATTACA CACTTAGGAG CATAAATACC 113851 ACCACCACTT CTTTTTGCCG TATTGTTAAT GAAACTTAAA CTCTCATTTT 113901 CAGTAAGAGT TAAGCTTTTT GTAGCTATGT CAGACTCTGA GATATTACAG 113951 AGGATCGCTC CACCACAACC TTCTTGATCT GTAGTTGTTG TTGCTGTTGC 114001 TGTTGCTGAA TTTCCAGAAA ATACAAGAGC CTTATTTTTG GTAAAGGAAG 114051 TATTTCCTTT AGTATGTAGA GCCCCTGCTG TCTTTGCTGT ATTTGTGCTG 114101 AAGGTCACGG TTCCCGTACT TCCCGTAAGA GTAAAATCTT CGGTTTCTGT 114151 ATAGATCGCT CCTCCTGCAG TTTCGGCAGT ATTGCCATCA AAGGTAAGAG 114201 TCGTGTTTCC ATGCAGAGCA CACTTGGTCG CATAGATCGC ACCGCCACTT 114251 ACTGTTGCAG TATTACCAGA GAGACTCACG TTTTCGTTAT CTTCAATCCA 114301 GAGTCCTTTT TTAGTACTTA CAGATGCTGA CTCAAGAAAC GATAGGATTG 114351 CCCCACCGCA ACCCTCTTGA TTTGCTGAAG AATTACTCGG GCCCGTAGCT 114401 TTGTTCCCTG AAAAGAGCAG GTTGGTATTA CCAGACAGAG AGTTGTTGCC 114451 TTTAGAATAT AAGGCGCCGC CTGTCTTTGC TGTATTTGTG CTGAAGGTCA 114501 CGGTTCCTGT ACTTCCTGTA AGAGTAAAAT CTTCAGTTTC TGTATAGATC 114551 GCCCCTCCTG AAGTTCCAGC AGTATTGCCG TCAAAGGTCA GGGAGCCGTT 114601 TCCAGTTAGA GTACATTTGG TAGCATAGAT CGCACCACCA CTTACTGTTG 114651 CAGCATTACT AGTGAGGCTG ACTTCTTGGT TGTTTGCAAT CGATAGTCCT

114701 GTTTTATCGC TTACGGATCC TGAATCAATA AAGGCTAGGA TTGCCCCACC 114751 GCAACCCTCT TGATTTGCTG AAGAATTACT CGGGCCCGTA GCTTTGTTCC 114801 CTGAAAAGAG CAGGTTGGTA TTTCCAGTCA GCGAGCTGTT TCCTTTAGAA 114851 TATAAGGCGC CGCCTGTCTT TGCTGTATTT GTGCTGAAGG TCACGGTTCC 114901 CGTACTTCCC TTAAGAGAAA AATCTTCAGT TTCTGTATAG ATAGCTCCGC 114951 CACATCCTGC TGTCGCAGTA TTCTGATCGA AGGTAAGAGT TGTGTTTCCA 115001 TCCAGAGTAC ATTTAGTAGC GTAGATCGCT CCACCATTCG CAGTTGTTGT 115051 ATTACTAGTG AAGCTCATTT CTTGATTCTG AGAAATGGCT AATCCAGTTT 115101 TGTCTGTTGC TGTAGCAAGA TAACAACAGA TTGCCCCACC ACAACCTTCC 115151 GGGTTATTTG CCTGTGCTGC TGAGCCGGTT GTTTTATTTT CCTGAAAAAG 115201 TACTTGAGTG TTGCCGGTAA GAGCAAGATT GTCATCAGAG CTCCAAGCAC 115251 CCCCCGTCTT TGCAGTATTA GATTTGAAGG TAACGACTCC TGTATTGGCA 115301 TCTAGCGTGC TATCCTTTTC TTTTGAGTAG ATCCCCCCAC CTTTATCTGT 115351 AGCAGTATTT GAGGAGAAGG TCACCGTTCC TGAGTTTCCT TGGACTGTAG 115401 TGTTTGCTGT ACTACAGAGG GCCCCGCCAT TTTTTGTGCT AGTATTTTGA 115451 TCTAAGAGAG CTGCTGTCGT AGTCTTAGCA AGATCGATGC TGTAGGCAGA 115501 AACTGCAGCT CCATCTTTT CTGAAGTATT TTTTTGGAGG GTGACACTGG 115551 CATTGTCAGT AAAAGTCGCA GTACCTCCCT CTGTATTTGT CACACAAATA 115601 GCACCCTTGC CGCCCGAAGT TCCTGTTGCT GGAGCTGAGT CGATTAAGAG 115651 TGACGAGAAT CCTGAGAAAG AAAGAGCTGT GTTGGTATTG TTAATTGCAG 115701 CACCATCATG CGTAAGCGCT ATGGTTTGCA GAACCAATGA GTGATCAGCT 115751 CCAACAAAAC TCAATGCTCC TCCTGTGTTT GTAAAACAGC TTTTATCTGC 115801 AGGAGTAATT GCAGATACAT TCGTAATAGA AACATCGCTA GTGAGAGTGT 115851 AGGTAGTTCC TGAAGCATCC GAAGTTTCCT TGGCAGTGAA TGCTGCGCTA 115901 CCACTACTAC CATTTCATA GTTATCGGAT GATGAGAGAT CCGTGTTAGC 115951 AGCCATTAGT GGATGTAGGG AGAAAACTAA AGCCGAAGAG GTAAGTAGCC 116001 AAGGTAAAGA ATATTTCATG TGTCTTTGGG GAAAAGCTTT TTATCAAAAA

116051 TACTCCCATA GCATGTGGCT TTAGGAGCAT GGTGCACCAA TAGAGAATAC 116101 AGTTAAATAA ATCAAGTAAA TGCTCTGGAG AAGACTCTCA GTTATAGAAG 116151 TTTCAATCTT GGGAGAGAG CATTTAAGGT ATTTTCTAT ATTTAAGAGT 116201 CCCTTAAAAC ATAAGGGAAA TGCTTAAGGG TAGGGGAGAA GGTGTACAAG 116251 CGGTTTTGCT TTTAGACCTT CTTGAATTTT AGAAGGAGA AGTAAGGAAG 116301 ATGGTTATCT AAACCACGGA TCCTTATTTG TTCTTTCTGT GTTTCCGTTT 116351 TTCTCTTTCG TCATAACCGT CAGAGAATGG ATTGGAGGGG CTGAGGTTAG 116401 CCGTGCTTTT TCCGTCTGGA GGCGGAGTTT TAAGAGATTG ATTCGATTTC 116451 CCGCTTTTCT TCTTAGATTG CTGTTTCTGT TCTTCATCTT GATTTTGCTG 116501 TTGCTGCTGT TTATCTTGGC GATCACGAGG GGAACGCCGG GAAGATGAAG 116551 AAACTCGAGG AATTGAGGGT TCTTTTCCTC CCTTAGGGTA GACCGGCTCA 116601 GGGTGTAGAA CCGTCCCCGT GGAGAAATTA GGAGAGCGGG AACTTGCAGG 116651 CATTATGGGT GTAAACGCAC TGCTCGCTCC ACTACCGAAT GAACTGCTTC 116701 TTAAATCTTT GAAGTGATAA GGCTGAAAAT CGTCCTTAAA TGGAGGACTT 116751 ATAGATGCGG GTCGTGTAGA AGCCGCATCT GAGGGTTTTG TTTTAAAACG 116801 TGAGAGGCTG CCGAAGGAGC GTCTATGCTC AGGATTTCGT GACGAATAGA 116851 AGAAAGATCC TTGAGCCTCC CGACCGATCC TACGATGACG GGCATCTTGG 116901 CGTTGTGAAG CTTCGTGTTC CTCCATGCGG GCAGATGCAG AAAGATCTTT 116951 TTTCTTTTCC ACTGCGATTT TTTTTGCATC GTCGGCGGAT GGAGTCACTA 117001 GAATAGGCTC CGCGGTCTCT TGGGTTTTCT TTTCCATCCT CAAGAGGTGT 117051 TCCCTACGGA AATAATCTAA GGTGAGCTTA TTCAAGGACA TGAAGTATAG 117101 CCCCACCGTA ATCAATCCCA TGACAAACAT AGGGTTGGCA AAGACTAACA 117151 TCGTCCCACT AGAGGCAACG AAAGCCCCTG CAATAAGACC CGCAGCAATC 117201 GCTAATACAA TGATAGGAAC GACGATAGCA GTGATTGCCT CACCGATTTT 117251 CTTGGCCTTC GGACTGTCCA GAATATCAGA AATAAGTAGA GTAACTCCCA 117301 AAGCTCCCAG GGCAAGTGCC GGAGCGAGAG CATAAAGCAT AAGGCTGTTT 117351 CCTGAAGCAG TCAGGAGAAT CGAAAGGATC GCAATTGCCG CAAGGGCAAT

117401 AATGCCTGTG TCATAGACAT AGCGCATTTT AGGATATTTA TCAGGAATAG 117451 TAATAAATCT TTTGAATAAC CCTGTGGATG ACGTTTTTAA ACGTTTTACC 117501 ACGCTTGGCT GCCCTGATGG GGATGCTGCT GCAGGAACTT GAGGTTGACC 117551 TAAAGGGTTT ATAGGGGGTT GGCTCATACT ATCAACTTAC TGTAATTATC 117601 ATTAGGCCCA TGAATTTTCA TTCATAGGAT ATATTTCATA CTATTATAAG 117651 ATTTAATAGG ATTTAGTTAG TTCTCTTTTC TTCTGAGTCT TAACTTTTTT 117701 ATTAAATAAA GTTTATTTGT TAAAATCTTA ACAGATTTTT AACTAAAACT 117751 TTAAGTTATT TTTATTTGGA ACTTTTAGTC GAAATAAGAC TCGCTTATGA 117801 GAGGGACATA CTCATCAGCA AATGGAGGGG GCGTGTGAGG TCGTGAGGGT 117851 GGAGGCTCAT ACGGGGGGAA AAGATTTTTA TGCACTTGGA AGAGGGTAAC 117901 GCTAGTAGTT ACAGACATGA GAAGCCCCAA GGAGATAAAG ACAACTGGAG 117951 GGGGGATAAA GACCAGACTA AAGACCAGAC CGATAATCAC AGCAATCGTA 118001 AGAATGTGGA GGAGCCAAGC CATAGCATAC GAAGCAAGAT GTTGGCAGCT 118051 TTTTGTTTTT ATAGCGTTAA AGAGAGCCCG TAGGGGTAGG GTCAAGGCAG 118101 CATATGTAGA AGCGCAAGGG ATCAGGATAA CCTTAAGAGC TCCTAAGACC 118151 GAGGATACCA GTACTTCTAT GGTAAGAGCA GTTCTGGGAT AGCTCTTCGC 118201 ACAGTTTTTT AATTTTCGAG CTAGTATTCG TTCCGGAAAA ATGCCATTCA 118251 GGTATAGCTG AGAGCCTTGT TTGCAGATAT TTTTGAATCC CATATCCGTT 118301 TGAAAGAGA TATTTTATGA AAAATTATGT AAAAATTCTA AGAGGATAGT 118351 GGTTTTTAGA CAATCGAAAT TCCTGAAAAG GCAGGAAAAT GAGAGACACA 118401 AGTAGACAAA ATCTCCTGAA GTTTTTGTAT GGGCCTGTAA AAAAATCTTT 118451 CTGGAAACTG GAAATTAGAA GTTCATTACA GCGGAACCAC CGAAAAATGC 118501 GGAGGTTTTA AAAGTATGGG ACGTTTTATT ATTGTTGTGA TCATCAGCTA 118551 TCGAGATATC ATTCCCAATA GACCAACCGT AAAATCCCTT AATAAAACTT 118601 CCCGGCCAAG GGGTGAGCTT CACGTTAGCT TCGATTTCAC GTCCTTGATA 118651 TTCAAAGATG CCGCGAGAAG AAATTAGGTG ATTTTTGTGA AGTTTTTTTC 118701 GGAATCTTGT AAGGCGGTAA GCAAGCCCTA AGTTACAGAC AGATGTCGAG

118751 CGGTAATCAA TAGAAAAATT CACAGGATAG ATGCAATTGA GAGTTAGTTG 118801 GTCGGTAGCC TTGTAACTAA CACCTACTAA AGGCCAAGCC TTCTCTTGAT 118851 GGAGGCCTGT TTCATTAATG ACGCCAAAAA TAGCAGAAAG CTTCTCAGTG 118901 GCCTGGTATT TTCCAGAAAG AACTCCTTGA TAGAGTCCAT AACCCATCTC 118951 AATATTTTTA GGATCCACAA GCCCAGAAAG AATGATAGAC CACTGCCAAT 119001 TTTTTAAGGA GAGTGTATAA GCTCCTAAAG AGAGGAGAAC ATAGTTATAA 119051 AAAGAAGTAT CTTGGAAAGT CGCCCACCCA AGTCCATTAG GATCTGTCTC 119101 CGAAATAGGA AGTGAGCTTT TCCATTGAAT ATCCGCACCT ATATAGCCAG 119151 TAGAAAACAG TAGCCCAGAA TGCTCTGTAA TCGGAAGTGT GCAGAGAAAC 119201 GTTCCATCGT ATTGACGATA GCCTATAGTT TGATGAGGCA GCTTTTTAAA 119251 TTTAGCATCG TTCACCTTTA GGTATTGTAC CTGAGCAGAG AAAGGACGTG 119301 GAGGAGGATT TTTACATGCT TCTTCATCAA TTCCACAAGC ATCTTGAACA 119351 ATAAAAATAG GAGTCGAGAG TACGTGTCCC GCAAATGCAG CGATGTGGAA 119401 GAGCAGTTTG AACATGTTCT GTAAGATTCT CCAACGTTAC TAGAGATTGA 119451 AATGGAATAT ACGTAATTTT TAAATTACTA TCTAATAATT TTCCTACTCA 119501 GGAGGGACTT TCAAGAAGAT TTCCCTAGAT TTAGGGGGAT GAAAGGACTA 119551 GAATTTTCT AATCAGCAAT CGAACAAAAT TACAGGGTTC TTGGAGAGGT 119601 GCTTAATACT ACGCATAAAT GATTTCTGAG ATGATTTCCG TTTCGGCTGT 119651 TTCTGTGCTA ACTTGTTTTG GCACTCAGTA ATTTTAAGCT CAAGATTTTG 119701 GATCGATTCC TTTTTCGCTT CAATTTCATT TAAAATTCTT GTTTTTCTT 119751 GATTTTGTGT TGTCAGCTCT GGAAGATGCA CGCTTTTAAT TGTAGAGAAC 119801 ATCAACCCGT ATCTCTGAAC AAATCCTGAT CCTATAGAGG AAATGTCTTG 119851 AGATATCTCA TCAATCAGTT TTGTTCCTGT CTGCTTATTT TTAAGAGGGA 119901 TTGCTAGAAG AAGAGCAAGA ATCAGTGTGA GTACGATAAT TCCCAAGCCG 119951 ATGCCACAGA TGATCCAGTT TCCAGTATAT CCTGCGTAAC CTGCTGAGAT 120001 CGTTCCCCCT AAGGTTAACA GTAGTGTTAA GGCAAAACAG AGCGTATGCT 120051 TTGTGGAAAG TTTGGAGCGA AAGGTCTCCC AAGGAGCCGC TGGAATCGGA

120101 TGAGGCTGCG TAATATAAGC ATTAGGAAGA TTGAGAACTT CTGTAGCATG 120151 AGAGAGAGG CTGCTCTCGG GGACTGGTGA TGGGGCTACG GAAGTTGCCA 120201 TGTTGTTTCC TCGAATCGTT GCTAACTAGT TTTGATTTGT CTTTTCATTC 120251 TTGTGAGAAA GCTCAAAACG TTTTTGATAA AGAGTTGATT GCAGTTTGAA 120301 CTCTTTTGT GAGAGGTCGC AGAGTTCTTC GTATAACTTT AGCGTATCTT 120351 GAGTGGTTTT TCTCTGTGCA TTGGTTACAA GCTGTAGAGA GGATTCAGGT 120401 CGAATCTTCT CCCTGATAAA TTGTACTGTC GTATGGAGCT GCTGAGGGAG 120451 CTGCCGGATG AATTTAATAA ATCCTACCAA GGCTTGTAGG CAGAGAAGAG 120501 TCAAAATAGT AAGAACAATG CCAACGGCAA TCAACAGAAT GCTTTGGCTA 120551 TAGCAACCCA AACAGATAAT AGCGATTCCA GCAAGAGCTA CAATCACTAA 120601 GATCGTAATC GCAGCAATAT GCATAGGAAT GGAGTGGGTG AGAAAAACAT 120651 TGCCCTTCTC CCCCAAAGCT ACGATCTCCT TATTCGTAAT GAATAAATCA 120701 GCAGACTCTT CCGGAAGGGA TGAGGGAAAT ACCCCGTTTA AAGTACTAGA 120751 CACAAAGAGA ACTCTATTAT TTGAGGAAAT AATTTAAGAA AAATGGTATT 120801 TTTAGTCAAT TAGTAAGCGA GTCATGCCTC TTAGTTATTC AAATTTTTAA 120851 AACCTTACCC TTCCTATGAG GAGACAAGTA AGAGAAATTA TGCAACAAAC 120901 TGTAATTGTA GCAATGTCAG GAGGCGTGGA TTCTTCTGTC GTTGCCTATT 120951 TATTCAAAAA ATTTACCAAT TATAAGGTTA TTGGCATCTT CATGAAGAAT 121001 TGGGAAGAG ATCGCGACGG CGGTCTCAGC TCGACTACTA AAGATTATGA 121051 TGATGTCGAG AGGGTCTGTC TTCAGCTCGA TATACAGTAT TACACCGTAT 121101 CTTTTGCTAA AGAATATAGA GAAAGAGTGT TCGCTCGTTT CCTCAAGGAA 121151 TACTCTTTAG GCTACACTCC TAACCCCGAC ATTCTTTGTA ACCGAGAAAT 121201 CAAATTTGAC CTTCTACAAA AGAAAGTCCA GGAACTTGGC GGAGATTACC 121251 TCGCTACAGG GCACTACTGC CGATTAAATA CCGAGCTCCA AGAAACCCAA 121301 CTCCTTAGAG GTTGCGATCC TCAAAAAGAT CAGAGCTATT TTTTATCAGG 121351 AACTCCTAAA AGTGCTCTTC ACAATGTGCT CTTTCCTCTT GGGGAAATGA 121401 ATAAGACTGA AGTTCGTGCG ATTGCAGCTC AAGCAGCTCT TCCCACAGCA

121451 GAAAAAAAG ATAGTACAGG CATTTGCTTT ATAGGGAAGC GCCCTTTTAA 121501 AGAGTTCCTA GAGAAGTTTC TTCCCAATAA AACAGGCAAC GTTATCGATT 121551 GGGATACCAA GGAAATTGTA GGGCAACATC AGGGAGCTCA CTATTATACT 121601 ATAGGGCAGC GGCGAGGACT TGATCTTGGA GGATCCGAGA AACCCTGTTA 121651 TGTTGTGGGA AAAAATATAG AGGAAAATAG CATTTATATT GTGAGGGGGG 121701 AAGACCATCC CCAGCTCTAC CTACGGGAAT TAACAGCTAG AGAGCTCAAT 121751 TGGTTTACCC CTCCTAAATC CGGATGTCAC TGTAGCGCTA AAGTCCGCTA 121801 CCGTTCTCCT GATGAAGCTT GCACGATAGA TTATAGCTCA GGTGACGAGG 121851 TCAAGGTGCG ATTTTCACAA CCCGTCAAGG CGGTAACTCC AGGACAAACA 121901 ATAGCGTTTT ATCAAGGAGA TACCTGCCTT GGTAGTGGAG TTATCGACGT 121951 TCCTATGATT CCAAGTGAGG GCTAGGGAGA GCAGCTTCCT GCTCCTCTTC 122001 TTCCCTTTCA AAGGCAACGC GATTTTCAAC CAAGGTTGCT CGTAGCTTGC 122051 GAGCTTCTTG ACGGCAGGAC TCTTTAAGCA ÁGAGCTCCGC TAGAGGATCT 122101 TCAAGGTACT GCTCAATGAC ACGGCGTAGA GGACGTGCTC CCATTTCTGG 122151 AGAATGCCCC TTCGTTACTA GGAAGGAAAT CACAGAGTCT GGGATGTTCA 122201 AAGCCATTTG GTAGTTTTTC AGTCTCGAGT CCAGTTTGTT GATCTCTAAA 122251 TGGATGATCT CCGATAGAGA TTCTTTCTCG AGGGGACGGA AAATCACACT 122301 TTCATCCAAA CGGTTAATGA ACTCAGGCTT TAAGTGTTTC TTCATAGCAT 122351 GTTCGATTTT CTCTTGGATG ACCTTATAGT CCATATGGGA CTTCAAGCCA 122401 AAACCAATTT CTCCGCTTTT ACGAATGAGA TCAGCTCCCA AATTGGAGGT 122451 CATGATAATA ATGGCATGAC GGAAATCCAC TTTGCGACCA AAAGAATCAG 122501 TAAGACGTCC TTGCTCTAAA ATTTGCAACA TCAGGTCCAT AATGTCTGGG 122551 TGTGCCTTTT CTATCTCATC AAAGAGAACA ACGCAGTAAG GACGGCGACG 122601 TACCTGTTCC GTAAGGTGGC CCCCTTCTTC ATGACCTACA TATCCTGGAG 122651 GTGATCCCAT CATCTTGGTA GCAGCAAATT TCTCCATGTA CTCTGACATG 122701 TCTACCTGAA TCAGAGCGTC TTCACCACCG AACATCTCTA TAGCAATTTG 122751 TTGGGCGAGC AGGCTTTTCC CTACACCGGT AGGCCCAAGG AATAGGAAGG

122801 AGCCCGTAGG TCGGTTAGGA TCTTTGATCC CTGTTCGAGA ACGTCGGATG 122851 GCACGGCAAA TGCTGGTAAC GGCATCATTT TGACCAATGA CTTTTCTTCT 122901 TAACGTGTCT TCTAACTTCA GAAGCTTCTC ACTTTCAGCT TCTGTGAGCC 122951 TTGCTGAGGG AATTCCTGTT TGTAGAGAAA CTACCTGAGC GACTGCTTCT 123001 TCATCTACAG GAACTTGGTG CTCTTCTTTA TGATTTTCCC ATTCCTGTTT 123051 CATACTTTGC AGACGTTCGC GAAGTTTTTT CTCTTCATCA CGTAAACCTG 123101 CAGCTTTTC GTATTCTTGA GTTCCAATGG CCTGCTCTTT GGCCAATTTT 123151 GTATTTTCGA TTTCAGCCTC TAGCTTCATT AAATCTGTAG GCTGACCCAT 123201 TGTATTCACA CGGACACGAG CCCCAGCTTC ATCTAAAAGA TCTATTGCTT 123251 TATCAGGGAG GAAACGTCCA TGAACATATT GATCAGAAAG AGTCGCAGCT 123301 GCTTTTAAAG CTTCTTCAGT AATGAAGACA TTGTGATGTT CTTCATACTT 123351 TTTCTTGAGG CCACGTAAAA TCTCAATAGT CTCATCTACA CTAGGAGGGT 123401 GAACCACGAT TTTTTGGAAA CGACGTTCTA AAGCTGCGTC TTTTTCTATG 123451 TGCTTGCGAT ACTCATCTAT CGTAGTTGCT CCAATACACT GAATTTCACC 123501 TCGCGCTAAC GCAGGTTTTA AAATGTTTGA AGCATCGATA GCACCTTCAG 123551 CTGCTCCTGC TCCTACAATC GTGTGGAGCT CGTCAATGAA GAGCAAGATG 123601 TTTCCATGCT TGCGAACTTC ATCCATGACA GCTTTGATCC GTTCCTCAAA 123651 TTGCCCTCGA TATTTTGTTC CAGCAATCAT TAATGCTAGA TCTAGAGTAA 123701 TCAGTCGCTT TTTCCGTAAG GCATCAGGAA CCTCATTCAG AATGATTTTT 123751 TGAGCCAGAC CCTCAACAAT TGCAGTCTTA CCAACTCCAG CTTCTCCAAT 123801 AAGTACAGGA TTGTTTTTC TTCTTCGGCA AAGAATCAAA ATCAACCGTT 123851 CGACTTCTGA AGAACGACCA ATGACAGGAT CGAGCTTAGA CTCTCGGACC 123901 ATCTCCGTTA 'AATCATAACC ATATGCTTTC AGAGCAGAAA GCTTTTCGTT 123951 TTTGTCAGAA CCTAAGCTAT GACCTAAAGG AGATTTTGAA GATGAAGGGT 124001 TGCTTCGAGA GGATGAGGAA GAAGACGACG ACGAAGGAGG AAGTTGTAGA 124051 TTGAAGGTCT CTAATTCTCT AAGAATTTCC TTACGAACCT CTCTTGGATC 124101 GATATGTAAG TTTTCTAATA CCTGAAGAGC GACACTATCT GATTGATGTA

124151 GGATCCCTAA GAGTAAATGC TCCGTCCCGA CATAATTGTG CTCTAAAAGG 124201 CTGGCCTCTT CATTTGCTGA TTCAAAAGAT TTTTTTACTC TTCCTGTAAG 124251 GGCAGGGTCT CCGTAGACTT GAATTTCTGG ACCATAACCA ATCAGGCGTT 124301 CCACCTCTTG CCGTGCCGTA TCAAAATCTA TACCGAGGTT GCGTAATACA 124351 TTAACAGCTA CCCCTTGACC AAGTTTGAGA AGACCAAGCA GGATGTGCTC 124401 AGTACCCAGG TAGTTATGAT TTAAACGCTG AGCCTCCTTT TTCGCCAGTT 124451 TAATGACTTG TTTTGCTCTA TTAGTGAACT TCTCAAACAT AAAAACCTAA 124501 AAGACAGGGG TAGAACTTTC CTTAAGCATA TACGAAATTT AAAATAATGA 124551 TGCAACTCTT CGCTCTAAAC CAGCAAATTT GGTAAAATTC CTCTGAGTTT 124601 AAGGGAAAGT TATGCACAAA CCTTTTGTAT ATGATACAAT AGTTCAGCTT 124651 CTTTTGAAAC AGTCTTAATT AGTTTTATGT TTGTTATATG AAAGTTCGTA 124701 TCGTAGATTC AGGAAAATCT TCAGCGGCCT CCCACATGGC TAAGGACAGA 124751 GATTTATTAG AATCTCTGCA AGATGGGGAG CTCATTTTAC ACCTTTATGA 124801 GTGGGAGAAT CCTTGTTCTC TGACGTACGG TCACTTTATG CGTCCAGAAA 124851 AATTTTTACT TTCCAACTAT GCGGATCTAG GATTGGACGC CGCAGTGCGG 124901 CCTACGGGAG GGGGATTTGT CTTCCATAAG GGAGATTATG CTTTTTCTGT 124951 TCTTATGTCT GCGACACATC CTTCCTATTC TTCTTCGGTA CTTGAGAACT 125001 ACCATACTGT AAACTCTTTT GTAGCGAAGG TTCTAGAGAA AGTATTTCGG 125051 ATCCAGGGAA TGTTAGCTCC AGAAGACGAA AACTCTTCTT CCAGAGATTC 125101 AGGAAATTTT TGTATGGCAA AAACTTCGAA GTATGACGTT CTTTTTGGGG 125151 ACAAGAAGAT AGGGGGCGCT GCCCAACGCA AGGTGCAACA GGGATTTTTA 125201 CATCAAGGAT CCTTATTCTT ATCGGGAAGT TCTTCTGAGT TTTACCAGAG 125251 ATTTTTAAAA CCCGAGGTTC TTGAAGAAAT TATTGAACAA ATCCAGATTC 125301 ACGCGTTTTT CCCTTTAGGT TTGGAAGCTG CTGATGAAGT GCTGCAGGAG 125351 GCGCGTCAGC AAGTCAAAGA GGCGTTTATT AAATTGTTTT GTGGTGAGGG 125401 GTTATGATGA GTCGGTTGCG TTTTCGCTTG GCAGCTCTTG GAATATTTTT 125451 TATTTTGCTG GTTCCTAATT CTGTTTCAGC AAAGACAATC GTAGCTTCAG

125501 ACAAGGAGAA GGTTGGAGTT CTTGTTTATG ACAATAGTGT AGAGGCCTTT 125551 CAACAGATAT TGGATTGCAT AGATCATGCA AATTTTTATG TAGAACTGTG 125601 TCCCTGCATG ACAGGAGGCC GAACGCTTAA AGAGATGGTA GATCACCTCG 125651 AGGCTCGTAT GGATCTGGTT CCAGAGCTCT GTAGCTATAT CATTATCCAA 125701 CCCACGTTTA CCGATGCTGA AGACCAAAAA TTACTCAAAG CTCTCAAAGA 125751 ACGTCATCCC AACCGGTTTT TCTACGTTTT TACAGGGTGC CCACCCTCAA 125801 CAAGCATCCT CGCTCCTAAT GTCATTGAAA TGCATATCAA ACTTTCTATC 125851 ATCGATGGGA AATATTGTAT TTTAGGTGGT ACCAATTTTG AAGAGTTTAT 125901 GTGCACTCCA GGGGATGAGG TTCCTGAGAA AGTGGATAAC CCACGTTTAT 125951 TTGTCAGTGG AGTGCGTCGG CCCCTAGCAT TTCGTGATCA GGATATCATG 126001 TTGCGTTCTA CAGCATTCGG TTTGCAGCTC AGAGAAGAAT ATCATAAGCA 126051 ATTTGCTATG TGGGACTACT ATGCACATCA TATGTGGTTC ATTGATAATC 126101 CTGAACAGTT TGCAGGCGCC TGTCCTCCAC TGACTTTAGA ACAAGCCGAG 126151 GAGACAGTAT TTCCTGGATT TGACAAACAT GAAGATCTTG TTCTTGTCGA 126201 CTCTTCCAAG ATCAGGATAG TTTTAGGTGG TCCCCACGAT AAGCAACCCA 126251 ATCCTGTGAC TCAAGAATAT TTGAAACTTA TCCAGGGAGC TAGATCTTCT 126301 GTGAAGCTTG CTCACATGTA TTTCATCCCT AAGGACGAGC TTTTAAATGC 126351 TCTTGTCGAC GTTTCTCATA ATCACGGTGT TCATCTGAGT TTAATTACGA 126401 ACGGCTGTCA TGAATTAAGT CCTGCAATTA CAGGACCCTA TGCTTGGGGA 126451 AACCGTATTA ACTATTTCGC CTTGCTCTAT GGGAAACGGT ATCCTCTTTG 126501 GAAAAAATGG TTTTGCGAAA AGCTAAAACC TTATGAGCGG GTTTCTATTT 126551 ATGAGTTTGC TATTTGGGAA ACGCAGTTGC ACAAGAAGTG TATGATTATC 126601 GATGATGAAA TTTTTGTGAT CGGAAGTTAT AATTTTGGAA AGAAAAGTGA 126651 TGCCTTTGAT TACGAAAGTA TTGTAGTTAT CGAATCTCCA GAAGTCGCTG 126701 CAAAAGCTAA CAAAGTCTTC AATAAAGATA TCGGATTGTC GATTCCTGTA 126751 AGTCATGGCG ACATTTTCTC TTGGTATTTC CATTCCGTAC ACCACACTTT 126801 GGGACATTTG CAGCTGACCT ATATGCCAGC CTAGCGTCCC TGGGTGCGAA

126851 TCTACCAACA GGATCTCTTC TGCAGGCTCT GCAGGGATCC TGCCTGGTTT 126901 TTTTCTCTGC TATCGTTTAC ACTACGCTTT TATTGTTTGG GTAGAGGGTG 126951 GACCTTGTTA TCGTTCTTCT ATAAGCATCA AAAAAAATTT ATCGGCATTG 127001 TCATTGCTGT AGTTTGTGTT TCTTGGTATT GGAGTGGGTT GGGGACGATT 127051 CTCTAGAAAA GGTTCTGCAG AGTCCACCTC ACGTCGGACT GTTTTTACTA 127101 CCGCTTCAGG GAAGCGGTAT GTAGAGAAAG ATTTCATGGC TATGAAGAAG 127151 TTCTTTGCTC ACGAAGCGTA TCCATTTACA GGGAACCCTA GAGCTTGGAA 127201 TTTTATCAAT GAGGGGCTAC TTACTGATTA TTTTCTAACG ACAAGGGTGG 127251 GAGAAAAACT CTTTTTAAAA GTGTACCATC CGGGAGAGAA AATTTTTAGT 127301 AAGGAGAAG CTTACCAGCC GTATCGTCGT TTTGACGCTC CTTTTATTTC 127351 CTCTGAAGAA GTTTGGAAAT CTTCAGCTCC CCAGCTTTTA GAGATCCTGA 127401 AGGTCTTTCA ACAAATCGAG AACCCCATAT CAAAAGAAGG ATTTCTTGCT 127451 AGAGCCAAGC TCTTTTTAGA AGAGAAGG TTCCCTCATT ATGTGCTTCG 127501 ACAAATGTTG GAGTACCGCA GGCAAATGTT TGCTCTTCCC CCAGATGAAG 127551 CCTTATCTCG CGGGAAAGAC TTGCGGTTAT TTGGCTACCA GACGATTCAA 127601 GACTGGTTTG GGGATGCCTA CCTTTCTGCT GCTGTTGAGC TCTTGATCCG 127651 CTTTATTGAC GAGCAGAAAA AAGTACTTCC CAGGCCCTCA AAACAAGAAG 127701 CTCGTGACGA CTTTTATGAT AAGGCGAAGC ATGCCTATAC TAAGATCAGT 127751 AAGAATAAGG AATTTTCCTT AGGATTTGAA GAATTTGTAA ACTCGTATTT 127801 TCAGTTTTTA GAGATCTCTG AGTCCGAATT TTTCAATATG TATCGAGACA 127851 TATTGTTGTG CAAAAGAGCT CTTCTCCTAT TGCAGGGAGG CGTTTCTTTT 127901 GACTTCCAAC CTCTAACTAC ATTTTTCGTT CAAGGAAAAG ATTCCATACA 127951 AGTAGAGTTC TTTAGACTCC CTAAGGAGTA TAGCTTTAAA ACAAAACAAG 128001 AGTTAAAAGC TTTCGAAGTC TATTTAAAGT TAGTGAGTTT ACCTAAATCG 128051 GATAGTTTGG ATGTTCCTAA TGAGATCCTT CCTATAGCGA CCATAAAAGC 128101 TAAAGAGCCT CGGTTAGTAG GCAGACGGTT TTCTATAGAC TATAAGAGAG 128151 TCGCTTTGCA AGACTTAGCA GCTACTGTAC CTATGGTTGA AGTGCTGCAC

128201 TGGCAACAAA ATTCTGAGCA CTTCCAGGAG ATTCTCCAGC AGTTTCCTGA 128251 CGTTGAGACG TGTCAGTCGT ATAAAGACTT CCAACATCTT AAGCCTGCGC 128301 TGCGAGATAA AATTTCTCTT TTCACACGCA AGGAAATCTT AAGGGCCCGC 128351 CCTGAGAGAA TTCTGCAATC GCTACAGCAA GTTCCTAAGC AGAGCCAAGA 128401 AGTTCTCTTA TCTGCAGGGA AGAATAGTGC TCTACCAGGA ATATCCGACG 128451 GTCAGCAATT AGCCAAAGTG TTGCTTGAAA ACGAGGTTTT AGATTTATAT 128501 AGCCAGGATG CAGAGACCTA TTATACTATT ATTGTTAATA GTTCTTTTGA 128551 AAAAGAAGAA GTGCTTCCTT ATCGTGAGGT TTTAAAGAGA GATTTGGCCT 128601 CACAGTTACT TACTTCTCAT GGTCATCTTG TTGACATGGA GCGTCTAGAA 128651 TCTGCGTTGC GTACACGGTA TCCAGGAGAA GAAGGCGCTA GCCTATGGCA 128701 ACGACGTCTT TGGAAGGTAG TGGAAAACCA CAGATTGGGA AGGCATCTCG 128751 AGGGGTCTTT CTCTTGGAGC TTAGATCGCT CATTGAAGAC TTTTTCCCGA 128801 GGAGACAAGG AGCTGCCCCA AGAGTTTGAT AGGATTTTCT CTATGAAGGT 128851 AGGAGACTAT TCTTCTGTAT TCATGAGTCC TAACGAAGGG CCCTGTTATT 128901 ATCAATGCCT CTCTCATTTA CTGTATGATC GTCCTGCTAG CGTGGATAAA 128951 CTATTTTAG CTAAAAGTCA GCTAGATGAA GAACTTTTAG GATCCTATAT 129001 GGAACGCTTT ATAGAACAGG GAGTCGTAAG GTGATGTGGT ATTCTGATTA 129051 TCATGTTTGG ATTTTGCCCG TCCATGAGAG GGTGGTGCGC CTCGGGTTAA 129101 CAGAAAAAT GCAGAAAAAT TTAGGAGCCA TTCTCCATGT GGATTTACCT 129151 TCAGTAGGGA GTCTATGTAA AGAAGGTGAG GTTTTAGTCA TTCTGGAATC 129201 TTCTAAATCT GCTATAGAGG TGTTAAGTCC TGTATCAGGA GAGGTTATCG 129251 ATATCAACCT TGATTTAGTG GATAATCCTC AGAAGATTAA CGAAGCTCCA 129301 GAAGGTGAGG GATGGTTGGC TGTAGTCCGA CTAGACCAGG ACTGGGATCC 129351 TTCTAATCTT TCTTTGATGG ATGAAGAGTA AATTTTTTAT TAGATATACT 129401 CATTTTTTC AGAAGATAAG AGGTATTTTT TTAAGGCTAA AACATTTAAA 129451 ATTTATGTCT AAGGTTTAAA AAATACATCA GAATTATTCT ATGGATCCAG 129501 CTAGTCCGGT AGCCCCTCAT GTCCTACAAG ATCATGTGCA ACTATCTTCT

129551 GAAGAATTGT CCGCATTATC TTCCGGGGTA TCTCGTGTGA AGAAGCTTAC 129601 TATAGCCATC ATGGTCCTTT CATTGATAGC GATTTCTTTG GTAGCCTGTG 129651 GCCTATTTTT AACGGGATCG GCACCTCTAC AGCTCTCGAT CTGGATTGCT 129701 GCGAGTTGCA TTACCTTATC TATGTTAGTT TGTGCGTGTT GGCGTTATAA 129751 GATTTCCAAT GCCTTAGAAA AAACTAAGGT AGCGCATGAA AGCTGAGTTG 129801 GACATTTTAT TGATTGAAAA ATCATGACTA CATTACCTAA GTACGTTCCC 129851 CGTTCTCGAC AAAATCCCGA TACTCTGACC TTCCTAAAAC GGTATTCTAG 129901 TGTCCTTCTC CATTCGGAGA ATTCTTTATC TTATCGGATT TTTGCGAAAG 129951 TGCTTGCTAT TCTCCTCACT TCGTTAGCTG TAGCTTTCGC CGTGACTTTG 130001 TTTTCTTGTG AAGGTTCTCA ACTGAGACTC TGCGCTCTCT ATATAGGTAT 130051 AGCTCTTGCT ATTTGTGTTT TACTGACGAT CGTTGTTTAT TGTATCGCAA 130101 GTAAAATCGC CACAGCTTGC AAAAAGCCGC CTTCCATATC TCGAATTGAA 130151 ATTGTTTAGA AGCATCTCTG TGTACAAAAG TTCACTAGAA ACTCGACTCT 130201 AGGAAAGTTC CTAGAGAGAG CGACGCTGCG TTCGTGCTTT AGAAATACTT 130251 GGCTGGTGTT TGGACGAAGA TTCTTTTAGT GGATTGGTCG TGTTTTCAAC 130301 AACTTCAGTC TCTAGAGGAG CTCTTTGAAT CTTTGCTGAA GGTTTAGAAA 130351 TATCAATACC TGTTAAGCTC ATAAAAGCCA TAGCAATGAG GCCTGTTGTA 130401 ATGAAGGAGA TCCCCATTCC CTGGAGGTTT TTGGGAATAT CAGAGTAGGC 130451 GAGTTTTCT TTGATAGTGG CTAAAATAAC AATAGCGAGC CACCACCCAC 130501 ATCCCGCTCC TAAAGAGAAG ATCATCATAG GAATAAAAGG ATAACTACGT 130551 GTGATTCCGA AGAGCACACC CCCTAGGATC GCGCAGTTCA CAGCAATCAA 130601 GGGAAGGAAG ATCCCTAAGG AGAGATATAG ATTCCTGGAG ACCTTTTCTA 130651 AAAGAAGCTC TAAGATTTGC GTGAATGCCG CAATCACCAC GATGAAAATA 130701 ATCAGCTCCA GAAAACCTAG GTTTACAGAA GCTAAAGATG GAGAGATCCA 130751 AGTTAGAGCT TTAGGGCCCG TGATGAAAGC ATGGACAAAC CAGTTGATGC 130801 TCCCTGTTAC AGTGAGAACA AGGGCTACGG ACATCCCCAA GCCATTGGCT 130851 GTAGAAACCC TAGTAGAGCA AGCAAGGTAA CTACACATCC CCAAGAAATT

130901 CGCAAGAAGG ATATTCTGAA TAAAGGCTGC TTGTAGAAGA ATACCAAAGA 130951 CATTAAGCCA AGTATACGCA CCTAACCACA TAAACTACCT TTTTCTCTTT 131001 TTAGAGTCTC GAATGTTAAC AAGCCAAATC ATAATACCAA GTAGGAAAAA 131051 AGCCGACGGT GCTAGCACCA TAAGACTTAA ATTTTGGTAT CCATCGGGGT 131101 GGGTTTCGGA AGCATAAACA AATTGAGGGA TGATGCGAAA CCCCATAAGA 131151 GTTCCAAAAC CAAAGAGTTC TCTGATGACT CCAATGACAA GTAAGACCCA 131201 GCCGTATCCT AAGCCAGAGG CAAACCCATC TAAGAACGCT GGAATAGGAG 131251 TCACATGCCT AGCTAGACTT TCAGACCTTC CCATCACGAT GCAATTGGTG 131301 ATGATAAGAC CCACAAAAAC AGAAAGTGTT TTGGAAATAT CAAAGAAAAA 131351 AGCTTTTAAA AACTGGTCGA TAACAATCAC AAACAAGCTA ATGATAATTA 131401 GCTGAGTAAT CATTCTCACA CTGTCAGGAG TGAACTTACG TAATAAGGAA 131451 ACAAAGAAAG ACGAGCATCC TGTAACAATG CTGACAGCAA TTCCCATAGT 131501 AATTGCCGTT TGTACTGTTG TTGTCACTGC CAGAGCCGAG CAAATCCCCA 131551 AAATCGCAAT GAGAATTTGG TTGTTGCTCC ATAGAGGATC AAAGAAATAG 131601 CTTTTATAGG ACTTTTTACT TGTCATTCGC CTGTTTTCTT TTCATGGGTT 131651 AAATTAGAAA AATTTATAAG GAGCTGACGA TAGCAAGCCA GAGATTGTAC 131701 ATAAGCTTCA GTGACACCGT TGCATGTTAA GGTGGCTCCA GAAATCCCAT 131751 CAATAGCAGA AAGAGCTTTT GGAGAATCTC CCAAAGTAGT ACGCACGGAA 131801 CCTTTAACTA CCTCAAGCCC TAGGTCTGTT GTTGCAAAAT TTGTAGTTCC 131851 AGAAGAATCT TGTAGGAAGA TTTTCTTCCC ATAGAATTGC TCTTGCCATT 131901 CGGGATTTGT AATATTTGCT CCTAAACCTG GAGTTTCTCC TTGTTGGTAC 131951 CATGCGGTTC CCAATACAGT GTCACCGTCG TTTTTCACTC CTAGATAGCC 132001 ATGGATGGGG .CCCCAAAGGC CGAATCCTGA TATAGGGAAG ATCAAAGCTT 132051 GAACTGTAGA AAGGTCTTTC GCAACGTCGG CTCCTGACAT ATTTTCTGTG 132101 CGAGAGGTAT TCTCTAAAAT GACATAAAAG GGGAGGGGGG ATTGCTGACA 132151 CGGAGGGCTT TCTTGATATT TCTCAAAAAA TTCAATGGGA TTCAGATTTT 132201 TTTCTTCAAA AGAAAATACC TTGCCTTGGG CATCTGTAAG TAGAGGACGG

132251 ACAAAGCGCT CGGCATACAG CTCTAATTCA GGATAGGAAA CCTCAGAGAC 132301 TTTTTTGTA GCAACTTCAA GAAGTTGTGT TTTTTTATCG AAAGTCGCAG 132351 GCACCCACTC TTTTTTTCC TGAATTTGAA ATCTTCCTTT AAAATCTAAA 132401 ATATGAGCAG CTAAAAGCAT TTGCTTATTG CGATCGAAAG TAGCAGCTTG 132451 TTCCTGTATT GGGGAGAGCA CATAGTAGAT TGTGGATAAC AGCACTCCTG 132551 ATGCGGACGG TATGTTTTGA AGAGCCTTTA GACATATTCT AGACTCCCCT 132601 TTTTCTATAC TTTCTAACAG CAAAATAGTC GATAAGAGGG GCAAATACAT 132651 TGCCCAGAAG GATCGCTAAC ATCACTCCCT CAGGATACGC AGGATTGATA 132701 AGACGAATCA CAATAGTCAT AAATCCTATA AAGAATCCGT AAATCCATTT 132751 CCCTAATTTC ATAGTCGGCG ATGATACGGG ATCCGTAGCC ATAAAGACTA 132801 AACCAAAAGC AAGTCCTCCG AGGAAAAGCT GCCGATAGGC GGGAATGAAG 132851 AATCGAGCAG GTGCCCAAGC TCCGTTTTGT CCCACGATGA GTACGCTGAT 132901 AAACTTAAAG AGCCAGCCTG TGAGAAAGGC TCCTATCCCA AAGGCTGCCA 132951 TGGTTCTCCA AGAGGCAATG CCTGTAACAA TAAGGAATAT TGCACCCAAC 133001 AGACAGGCGA AAGTGGAGGT CTCCCCCAGA GAACCTATAA TGTTTCCCCA 133051 AAAGAGATTC CCAGCTGAGA ACTTCCCAAT CCCATAGATC ACATCGGTAA 133101 TAGCATAGGC AGAATCGAAC TGTGTGGGAA GCAGCCCCAA TCCTCCCTCA 133151 GCAACAGGAG CTGTAACAAA CGTTTGAAGT TGTGTAAGAG TGAGATTATC 133201 TAAAACCCAA CCAGGATGCG TCTCTGTCCA AAGAGAAAAT TGTGAGTGAA 133251 TGACATCTTG AGTAGGGACG TGAGGAATGT GAAGCATATT TGCAGCAATC 133301 GCATCGACAT GCAGACGCTT TACAGAGGGA GGTGTCGAAT TTAGAGTTTG 133351 TAGGCAGGTA GACTGTGAAA ATCCATCAAT GAGTACTTTT CCTGTCGAGG 133401 AGTTCATCTT CATGAGGCTA TCTTTAATCA CTCCGGGGTT GCTTCCTACC ·133451 CAAACGTCAC CACTCATCTT TGCTGGAAAC GTAAAAAATA AGAATGCCCT 133501 TCCTGATAGA GCAGGATTGA GGATGTTCAT CCCTGTGCCT CCGAAGAGCT 133551 CTTTACTGAC AACAATACCA AAGGCGATCC CTAAGGCTGC CATCCAGTAA

133601 GGAATTGTCG GAGGGAGAGT AAGGGGATAG AGGATTCCGG TTACTAGCAG 133651 TCCTTCTGCG ATTTTATGCC CACGAACTAC AGCAAATAGG ACCTCACAAG 133701 TACCCCGAC AACATAGCTA ATCGTAAGTA GAGGAATAAA GATCTTAAGT 133751 CCTTCCCAAA GGATAGGAAC TATATGGATC TCTTTGTAAA CAAAGGATAA 133801 ATAACTACCA AATCCAGAAA TATGTAAGAA TTGCTCCATC AGCACAGGAT 133851 TGCCTGAGCT ATAAACGATA GATTGAAGTC CTGAATTCCA GATCGCAACA 133901 AAGGTCGCGG GAAACAAGC GATAACAACA AGCATCATCC AACGCTTAAC 133951 ATCTACAGAA TCGCGGATGA AAGGAGGCTT GGAAGGGGTT TCAATAGGTT 134001 CGTAACAAAA TGTATCTATC GCATCGACAA TGGGAGTAAA GCGCTGATAC 134051 TTGTCTTGTT GACATAGTTT CCAAAGAGAA TTTATGAATT TTTTGAGCAT 134101 TGTGATTGAG AGAAAAGTTG AAGCTTCTAC ACGTTCAAAA ACGTAGCAAA 134151 CTGCTTAAAA TTTTAGGAAT AAAAATTTTC ATGATTCAAA AATAGATGGT 134201 ACTITITICG TIGCTGTTIC CAAAGITATG TIATGGCTGT CAAGCTCCAG 134251 GAGCCTACTT TTGTTCCAAC TGCTTGGAAA AACTTCTCGT AGAAGATAGA 134301 GAAGGGCGTT GTCTACATTG TTTTCGTTAT CTTGGTTCTT CCGAAACACG 134351 TCTATGTAGC CAGTGTTCAC CCTCTTCACA ACTTCAAGCT TTCAGCTTGT 134401 ACCTTCCTTC GCAAACGCC CTCTCGGTAT ATGCTCGTGC TTGTGAAGGT 134451 AAGCGACCCG CTCTGCAGTT TTTTTCTAAG AGTATCGCCT TTGAGCTAGC 134501 TTCACTGGAT GAGACTCCGA GTTGTATTGC CTATATAACA TCGACAATTT 134551 CTAGGAAAAT CGTAGTAGAA GTTGCTAAAC TAGAAAAGCT TTTACGCATT 134601 CCCTTGTGGC CGTGGCTTCC TAAGAAAAGA CAAATAGAAA AACTTCCTAA 134651 AGGGGAAGGT ATCTGCTTTT TGTCGGCCTA TCCTTTATCA CAAAAATGGA 134701 TGCAAACTAT CGTTGGAGGG AGTGCATCAC CTCTAGTATC TATAAGTCTC 134751 TTTCTCTCTC AGAATGATCA GTAATTCCTG CAATTGCAAG GTAACCAAGA 134801 ATACGTACGC CCTCATCAAC ACACAACGTG TTCGCTTGGA TGTCTCCTTT 134851 AATGATTGCG CCTCCACGGA GTTCGACTTT TCCAGATACT GTGATATTTC 134901 CTTCTACAAC CCCTTCAATA ATGGCTTCTT GTAGCTGAAT ATCTGCCTTT

134951 ACCACTCCTT TAGGACCGAT AATAATTTTT CCTTTTGAGA CTAAAATGCC 135001 TTCAAAAGTT CCGTCAATAC GTAGGAGACG TTCAAAAGCA AGTTCTCCTT 135051 TAAAGGTGAC GCCTTCTCCT AAGGTAGTTT CAGGTTCTTC AAGAGGGAGT 135101 AGAGATTCTG TTCTTGGAGT TGAGGACCAT TGAGGAAGAG AAGATTCTTC 135151 AGTTAAATTG TGATTCAAAG GGCGAGCTTC CGAAGCTTTA GGGTTGTCAA 135201 AAAGACTTGG AGGGGTCTCT GGGCGCTCGG ATCTTGAATA TGGCGAGTAG 135251 CTGGAAGGTG AAGAAGTTTC TTCTTCGTAA AGTGTTTGCA CATCTTCAAA 135301 AGGACCTTTT CCTGTTCTAC GGAACATGGG ACACCCCCTA AATTAACCAA 135351 CAATATGATT TTTACAGAGA TTTACTTGAC GCTTAAGAGT TTCGTCATTT 135401 TGTAATTTAT GTTCTATAGT TTTACAGGCA TAAAGTACTG TCGAATGAGT 135451 TTTACCAAAA GCAGCTCCTA TTGCAACTAA AGAATCTGTA ATAAGAGTTT 135501 TTGCTAAATA CATAGCAATT TGCCGAGCTA ACACAAGATC TTTAGAGCGT 135551 GAGTTTCCCT TAAGATCATT CAGCTTTACT TGGAATACTG TAGCAACACT 135601 TTTTAAGATC GTTTCTACAG AAATTTTTTG TTTTGTTGGA GAACGGAAGA 135651 GCTCTTTTAG AGTTTCTCGG ACTGTAGTTT CTGTAAGAGA CTTGCCGAAA 135701 AGACGACAAT AGGCAGTCAG CTTGTTGATA GCTCCTTCCA ATTGACGGAC 135751 ATTGCCATAG ATGTGATCCG CAATATAAAA TGCCATTTCA TTAGGAATGA 135801 GCAATCCTTT TTGCTCCGCC TTGTGCTGTA AAATCGCAAC CCGAGTTTCT 135851 AAATCAGGGA TGCCGACGTG AGCAACCAGT CCCCATTCCA TTCTAGCAAT 135901 GATACGCTCG GAAAGTTTGA GCTGACTTGG AGGTTTATCA CTGGTAATTA 135951 CAATTTGCTT ACTCAGGTTG ATCAAAGTCT CAAAGGTATT GCAAAACTCT 136001 TCTTCAAAAT TTTGGCGATT CTGTAAAAAT TGAATATCAT CAACAAGAAG 136051 TAAATCTAGG GAACGATAAA AATTTTTCAT TTTATCAACA GACTTGGATT 136101 TGAGATGGTA GACAAGATCG TTGATAAACG CTTCTGTAGT GATGCAATGG 136151 ATGCGTAGAT TTTTATGATG TTCTCTTACG TAGTGACCTA CGGCATGAAG 136201 TAAATGCGTT TTGCCTAATC CCACACCCCC ATGGATGAAT AAAGGGTTGT 136251 AGGAGCGGCC AGGTTTCCCA GCAATACCTA CAGCTGCAGA CTTCACAAAT

136301 TGATTTGAGG GACCTTCAAT GAAATTATCA AAGCGATAGG AGAGATTCAG 136351 CTTTAATTCA AAATCTTTAG TTTCTTCAAA GACCTCAGAA ATTCCTTCGT 136401 TTGATTCTTT TTGAGAAGCC ACGGGGGCTG AAGGTTTCTT GTGTTCTGCA 136451 ACTACAAATT CTAAAGCAGG CTCTCCATGA ACATCTAAGG GGACAAAAGA 136501 ACAGAGGTCT CTTTTGTAGT TATCAAGAAG ATAATTTTGT ACAAAAATGT 136551 TGGGGACTTC TAAGCGAATT TTCTCTTGAG TTTCTTCAAG AACTTGAATA 136601 GGAGAAATCC AATTTTCAAA AGCCGTTTTC GAGCAACGTG TCTTAACATA 136651 ATTTAAAAAC TGTTCCCAAG TAGTGCACTC GTTACAGGTT AACATGCCGC 136701 TCTCTTTATT TATAAAGCTT TCCCAAATAC AATCGACCCA TCCCATGAGT 136751 GATGGCGAGA AAATCTCATT GCATCTTGAC TATTCATTGC ACTGAGTGCA 136801 ATGACTCTTC GCGTTCAGAT TGGATCCTGC ATTCCCTGGG TGAAGTTCCT 136851 GTTAGGAATG GATCCTATAC ACCCTTCCTC CGCAGGTAAA CGCGGTACGC 136901 TCTGCCGATA AAATAACTCT ACCGATTACT AGGTTTTAAG GCAAATTGGA 136951 TCGTTGGTTT CGTTAGGCAA TAAGGAACCA CAAATTCAGG AAAAAATAAT 137001 TATGAAATTT TGTAATAAAA ATGGAAAAAG AACTAAAGAA ATCCGGAGTT 137051 CTTCAATCAC GAAATACGTT TTCTATAGGA GAAAAAATTA ACGAACTAAC 137101 GCAGCATTTT TTTTGGTTGC TTTACTATAA CTCATCAATA GAGCTTCAGC 137151 ATCATTAGCA ATTGCTGGTA TGGGACAGGA TGAAAGGTAG GTGGCAATGG 137201 CAGTAGCTTC TTCAATTCGT CCCAAACAGA AGAGAGCTTT TGTTTTATTT 137251 AAGAGTGTAG GCAGATGATC TCCTTGCATG CGGAGTGCCT GATCTAAAAC 137301 AGCAAGCGCC TGACTATTTT CACCAATTTG GAGATAAAGA CCTCCAAGAG 137351 TTTGATGATC ATAGATACTT AAAGGATCTA AGATCACTAG AGCTTCAAAA 137401 AAAAGAATCG CTTTTGAATA ATGCCCTTGG CGTAGAAAAG AATATCCTGA 137451 GATTCTGAGT TCTTCTAACT CATCATCTCC CCAGCCTAAG ATTGCTTTCC 137501 ATTCATTATC CAACATACAT TATCCTTGAA CAAATTGAAA GATACGAGAG 137551 ATCACATAGT CTATCATCAC ATTTGTAGAC TTGACTCCAA GATCTAGAGC 137601 TTTTAAGAGC ACTTTCCCTT GTTGTACTCT AGGATCGTCA TCGTCTTCTT

137651 CAGGATCTTC GTCTTCCTCT TCTTCTTTAT CTTTGTAAGA CAAAAAGGGA 137701 GTAATTTGAC TGCGATAGGA AAACTTCCCT CGAGTGAGAA CTTTTAAAAA 137751 TGAGGAGATT TTTTCTATGT CTTCATCTTG TTGGTCTGGA GATCCTAAAG 137801 AAGGTGCCAG GTAGGGTGTG GAAAAACGCT GTTTGTAAAA ATTATTTGGA 137851 GGGGAAAAAC ATGCCCAGTG CGACTTTTGA TTTGTCTGTA AAAGAGACGT 137901 CAAAGCCGAA GGCTTGGGGT TCATATCCAA AATTTGCGCA TGCTTGGCAA 137951 CATCACGAAT GGAGATGCCT TCCATCTGGA TTTCTTTGCG AAAGTCGCTG 138001 ACTATCCTAT TATTGGAAGC ATGTTGCTCA TATATAGACG TGCTATAATT 138051 AAAAATTTCT ACCATGGCGA GGCCTTAAAA AACCGTCTCT ATCTCTAGAT 138101 GATAGTGCGG TCTAGGAAAA AGTCAATAGT CTTGATCAGA AGCCTGAACC 138151 TTTTCTCCTT ATCGAAGGCA CTTAAGAAAA AAGCTCACCC CTATCAAAAA 138201 ATTTAGAGTG GCGAACTAGC GAGAATTTAA GATAAGGGAA GGCTTTCTTC 138251 TTTCTATGGA ATCCTGTATC TTTGTCCTTA TTTGAAGCCA AGAGCTTAGG 138301 AAATTCTTAT GTCCGAACGT GCGCATATTC CCGTATTAGT TGAAGAATGT 138351 TTAGCTTTAT TTGCTCAACG TCCTCCACAG ACTTTTCGAG ATGTCACCTT 138401 AGGAGCTGGA GGACATGCGT ATGCTTTTCT TGAGGCGTAT CCCTCTCTAA 138451 CTTGTTATGA TGGCTCCGAT CGAGATCTTC AGGCTTTGGC AATTGCAGAA 138501 AAACGTTTGG AGACCTTTCA AGATAGAGTC TCCTTTTCCC ACGCCTCTTT 138551 TGAAGATCTT GCGAACCAAC CCACTCCACG TCTTTATGAC GGAGTTCTTG 138601 CAGATTTAGG AGTCTCTTCT ATGCAGCTGG ATACTCTATC CCGAGGGTTT 138651 AGCTTTCAAG GGGAAAAAGA AGAGTTGGAT ATGCGTATGG ATCAAACGCA 138701 AGAGCTTTCC GCTAGCGATG TCCTGAACTC CCTAAAAGAA GAAGAACTAG 138751 GGAGAATTTT TCGTGAATAT GGAGAGGAAC CACAATGGAA ATCTGCAGCT 138801 AAAGCTGTTG TCCATTTTCG TAAGCATAAA AAAATTCTTT CGATCCAGGA 138851 TGTAAAAGAA GCTCTTCTTG GCGTTTTCCC TCACTATCGT TTTCATAGAA 138901 AAATACATCC ACTCACCTTG ATTTTTCAAG CTCTACGTGT TTATGTGAAT 138951 GGAGAGGATA GACAATTGAA AAGTTTACTA ACATCTGCTA TATCTTGGCT

139001 GGCTCCTCAG GGACGCTTG TCATTATTTC TTTTTGTAGC TCTGAGGATC 139051 GTCCTGTGAA GTGGTTTTTT AAAGAGGCGG AAGCTTCTGG CCTGGGGAAG 139101 GTAATCACAA AGAAAGTGAT CCAACCTACC TACCAAGAAG TACGAAGAAA 139151 TCCTAGATCG AGATCAGCAA AACTACGGTG TTTTGAAAAA GCTTCCCAAT 139201 GAACAAAGT CGTTTTTTAC GTTTATGCTG CTGTCTATGC TTTTGTGGAA 139251 GTCTCTTTTA TTTCTATATT AATAAGCAGA ACTCGCTGAC GAAATTACGC 139301 CTCGAAATTC CTTGTTTATC TGTACGCTTG CGTCAGCTTG AGCAGCAAAA 139351 TATTTCTTTA CGTTTTTTAA TTGATAAAAT AGAAAGACCT GATCATTTGA 139401 TGGAAATAGC AGCTCTTCCC GAATACCAAT ATTTGGAATA TCCCTCAGAA 139451 GAAAGTATCA GTCTTTTATC CTATGAGCTA CCGTAAACGT TCGACTCTAA 139501 TTGTTCTAGG AGTGTTTGCT CTTTATGCTC TTCTAGTATT GCGTTATTAT 139551 AAAATTCAAA TTTGTGAAGG AGACCACTGG GCCGCAGAAG CTCTCGGGCA 139601 ACACGAATTT TGTGTCCGTG ATCCTTTTCG AAGGGGCACC TTTTTTGCTA 139651 ACACGACAGT ACGTAAGGGA GACAAAGACC TTCAGCAGCC TTTCGCTGTC 139701 GATATTACAA AATTTCACCT TTGTGCAGAT CCTTTAGCTA TTCCCGAATG 139751 TCATCGTGAT GAGATCATCC AAGGGATTCT CCAATTTATT GAGGGGCAGA 139801 CCTACGACGA CCTCTCCCTA AAGTTAGATA AGAAATCTCG GTATTGTAAG 139851 CTGTATCCTT TATTAGATGT TTCTGTCCAT GACCGGCTAT CCCTTTGGTG 139901 GAAAGGATAT GCAACAAAGC ATCGCTTACC AACAAACGCC CTATTTTTTA 139951 TTACGGACTA CCAACGCTCG TATCCTTTTG GGAAGCTCCT TGGACAAGTT 140001 CTCCATACCT TAAGAGAAAT TAAGGATGAG AAAACAGGAA AAGCCTTTCC 140051 CACAGGCGGG ATGGAGGCGT ACTTTAATCA TATTCTGGAA GGGGACGTTG 140101 GAGAGAGAA GCTGTTGCGT TCTCCTTTGA ACCGTTTAGA TACGAATCGT 140151 GTTATCAAAC TGCCTAAAGA TGGCTCTGAT ATCTACCTTA CGATCAATCC 140201 TGTGATCCAG ACCATTGCAG AGGAAGAACT CGAACGGGGC GTGCTAGAAG 140251 CTAAAGCCCA GGGGGGTAGG CTCATTCTAA TGAACTCCCA AACAGGAGAG 140301 ATTCTTGCAC TGGCTCAATA TCCGTTTTTC GATCCCACAA ATTATAAGGA

140351 ATACTTCAAT AACAAAGAGC GCATCGAACA TACGAAGGTA TCTTTTGTGA 140401 GCGATGTTTT TGAACCCGGG TCGATCATGA AACCTTTGAC TGTGGCGATT 140451 GCTTTACAAG CTAACGAAGA GGCTAGCTTA AAATCGCAGA AAAAGATTTT 140501 TGATCCTGAA GAACCTATCG ATGTGACCAG GACACTCTTC CCTGGACGAA 140551 AAGGATCTCC GCTTAAGGAT ATTTCTAGAA ACTCTCAATT GAATATGTAC 140601 ATGGCTATCC AGAAATCTTC GAATGTCTAT GTAGCTCAGC TGGCTGACCG 140651 CATCATACAA TCTTTAGGAG TGGCCTGGTA CCAACAGAAG TTGCTAGCTC 140701 TGGGATTTGG AAGAAAAACA GGGATCGAGC TTCCCAGTGA GGCCTCTGGT 140751 TTGGTGCCTT CTCCCCATCG TTTCCATATT AATGGTTCCC TGGAATGGTC 140801 CTTATCTACT CCATATTCTT TGGCTATGGG ATATAATATT TTGGCAACAG 140851 GGATACAAAT GGTTCAAGCC TACGCTATCC TTGCAAACGG AGGTTATGCC 140901 GTCCGGCCCA CTTTAGTAAA AAAGATCGTC TCTGCTTCAG GAGAGGAATA 140951 TCATCTTCCT ACTAAAGAGA AGACACGACT CTTTTCAGAA GAAATTACTA 141001 GAGAAGTTGT TCGTGCCATG CGTTTTACAA CGTTACCCGG AGGTTCGGGA 141051 TTTCGAGCCT CTCCTAAGCA TCACTCTAGT GCTGGGAAAA CAGGAACTAC 141101 AGAAAAGATG ATTCATGGAA AATATGATAA ACGCCGTCAT ATTGCTTCTT 141151 TTATAGGTTT TACTCCCGTA GAGAGCTCGG AGGGAAATTT CCCACCTTTA 141201 GTGATGCTCG TCTCCATAGA TGATCCTGAA TATGGTTTGC GAGCCGACGG 141251 CACGAAAAT TATATGGGGG GGCGTTGTGC GGCACCCATT TTTTCTAGGG 141301 TTGCTGACCG CACACTCCTC TATTTAGGGA TTCTTCCAGA CAAGAAGCTA 141351 AGAAATTGCG ACGAAGAAGC TGCTGCATTA AAGCGTCTCT ATGAAGAATG 141401 GAATCGTTCT CCGAAACAAG GGGGAACGAG GTGAGGATCT CTATTTCCAT 141451 CTTGCTATAG ACTTTTACCG TTGAGCAAAG ACTCTCTATC AGAGAGCCCG 141501 TCTCCTCTTT ATCCTCTATG AGTAGTTTAT GTTATGGCTA GGGTAGGTCC 141551 TAAACTATAG AAATAACTTT AGCTTTCTTC CCCTAAATAA GAGACCAAAG 141601 TCTTGATGAG ACGGTCTATT GAAGTTTATG GAAGGGGGAG GTAAGGCTGT 141651 GTGTTTGGGG ATTTAGATTT GGGATAAAGG AGGCTTCTGT TCGTAGAAAC

141701 AGGAGAGCGA AATTTTATAT TTCAGAGAAG AGTAAGAACT TTATGGACAG 141751 TTTTTTGTGA TTGCTTGTAT ACTATCTTGA TTGAATTTTT TGTCGACCTA 141801 CGAGTAAAGA AATCCTTTAA GCATTTTTTA AAAATCAGAG TGAGAGCATG 141851 CCCCTAGAGG GCTTTTTATG AAAAAAGTTG TTTTTCAATA GTCCCTGGAG 141901 CGTAAATGGA TTTAAAAGAG TTACTCCATG GGGTTCAAGC TAAAATCTAT 141951 GGGAAAGTTC GCCCTCTTGA AGTGCGCAAC TTGACACGTG ATTCCCGTTG 142001 TGTGAGTGTT GGCGACATTT TTATAGCCCA TAAGGGACAG CGCTACGACG 142051 GAAATGATTT TGCTGTCGAT GCTTTAGCTA ATGGAGCAAT TGCCATTGCT 142101 TCTTCACTAT ACAATCCGTT TCTTTCCGTT GTTCAGATCA TCACTCCTAA 142151 TCTCGAAGAA TTAGAGGCTG AGCTTTCTGC AAAGTATTAC GAATACCCTT 142201 CAAGTAAGCT CCATACCATT GGGGTGACTG GAACCAATGG GAAAACTACA 142251 GTTACATGTT TGATTAAAGC TTTATTGGAT AGCTATCAAA AACCTTCAGG 142301 GCTTTTAGGA ACCATAGAGC ATATCTTAGG AGAGGGGGTG ATTAAAGATG 142351 GGTTTACTAC ACCTACACCC GCTCTTTTAC AGAAGTATTT AGCCACTATG 142401 GTACGTCAAA ATAGAGACGC TGTTGTTATG GAAGTCTCTT CTATAGGACT 142451 TGCCTCTGGA AGAGTAGCCT ATACCAATTT TGATACAGCA GTTCTGACTA 142501 ATATTACCTT AGATCATCTC GATTTTCATG GCACATTTGA AACCTATGTT 142551 GCGGCGAAAG CCAAGCTTTT CTCTCTCGTG CCCCCTTCGG GAATGGTTGT 142601 TATCAACACA GACTCTCCCT ACGCTTCTCA GTGTATTGAG AGTGCAAAGG 142651 CACCGGTCAT CACTTATGGT ATAGAGAGTG CTGCTGACTA CCGAGCCACC 142701 GATATCCAAC TTTCTTCCTC GGGAACAAAG TATACCTTGG TGTACGGGGA 142751 CCAAAAAATT GCGTGCTCTT CCTCATTTAT TGGAAAGTAC AACGTCTATA 142801 ACCTACTTGC TGCGATCTCT ACAGTACATG CAAGTTTGCG TTGCGATCTT 142851 GAAGATTTGC TAGAAAAGAT AGGCTTGTGT CAACCTCCTC CAGGTCGTTT 142901 GGATCCTGTA CTTATGGGTC CCTGCCCTGT ATATATTGAT TATGCACACA 142951 CCCCCGATGC TTTAGACAAT GTCTTAACAG GATTGCATGA GTTACTTCCT 143001 GAGGGGGAA GACTGATTGT TGTTTTTGGT TGCGGTGGAG ATAGAGATCG

143051	CAGTAAACGG AAGTTGATGG CCCAGGTGGT AGAGCGTTAT GGTTTTGCTG
143101	TTGTAACTTC AGATAACCCT AGGAGCGAGC CTCCTGAAGA TATTGTGAAT
143151	GAAATTTGTG ATGGGTTTTA TTCAAAAAAC TATTTCATCG AAATCGACAG
143201	AAAACAAGCA ATTACATATG CTCTGTCTAT TGCCTCAGAT AGAGATATAG
143251	TGTTAATAGC GGGAAAAGGG CATGAAGCTT ACCAAATATT TAAACACCAA
143301	ACAGTTGCGT TCGATGATAA GCAGACTGTT TGTGAGGTAC TCGCTTCCTA
143351	TGTCTAAGCA ACTGTCGTTT TTTGCTTTAT GTGTGTTAGG AAGTCACCCG
143401	ATTTTTGCTC AAACACCGAA TCCTCCTCAG CGTGTACGAC GCAGTGAGGT
143451	TATATTTATA GATCCTGGAC ACGGGGGAAA AGATCAAGGC ACGGCAAGTA
143501	AGGAACTTCA TTATGAAGAG AAGTCCCTGA CCCTGTCTCT TGCTTTGACG
143551	GTTCAAAGTT ACTTAAAGCG GATGGGTTAT AAACCTCAGC TAACCCGATC
143601	TTCTGATGTA TACGTTGACT TAGGGAAACG CGTTGCTTTG TCGAACCGTG
143651	GGCAGGGGGA TGTCTTTATC AGCATCCACT GTAATCATTC TTCAAACGCA
143701	GCAGCCTTTG GCACCGAAGT ATATTTTTAT AATGGTAAGG TCGGATCTCC
143751	GACTAGGAAT CGCATGTCAG AAGTACTGGG AAAAAACATT TTAGCTGCTA
143801	TGGAAAAAA TGGCATTTTG AAGTCTCGAG GTTTGAAAAC TGCGAACTTT
143851	GTTGTGATTA GAGATACTTC TATGCCTGCA GTTTTGGTGG AAACCGGGTT
143901	TTTATCCAAT AGTCGTGAAC GTGCGGCCCT GCAAGATGCT CGCTATCGTA
143951	TGCATGTAGC GAAAGGCATC GCCGAGGGAG TTCATAATTT TCTTTCTGGA
144001	CCTAGTTTTC AGAAACCAAA ACAGAATATC GCTAAAATAC GTAAACCACA
144051	GATACAAGCA AATTAGTACT TTAGGAGTTA AAGGCAAAAA ATCGTCCTCG
144101	ATGCGATTCG AACGCATGGC CTGCTGCTTA GGAGGCAACC GCTCTATCCT
144151	GCTGAGCTAC GAGGACGCAA AGACCAGCAC TTTACCAAGT TAGATTAAAG
144201	AAGTCACGTG TTAGATGCAC GCTAGCAATT TAGGGGAAGT TTTTCTCAAG
144251	ATGTGGGAAT GATTTTCTA GGTTCTAGAA ATATAGTTAT TCGCATTAAT
144301	CGATATGGTT TATAGTGATT GCGCATTTTT TAAAAAATGT CTTGAATCCA
144351	AAGGATGAAT AGATATGATG AGCTTAGACT TCAAAAGTTT ATTATTTACG

144401 GTTACTCAAC GAGCTACTTT AGGGCACTTT AATAGGAGGC ATTGTCTAAT 144451 ATGGCTACCA TGACAAAGAA GAAACTAATC AGCACGATCT CACAAGATCA 144501 CAAAATTCAT CCTAATCACG TACGTACCGT GATTCAGAAT TTTCTAGATA 144551 AAATGACCGA CGCCTTGGTT AAAGGTGACA GGCTTGAGTT TAGAGATTTT 144601 GGTGTGTTGC AAGTAGTAGA AAGAAAACCA AAGGTAGGAC GTAATCCTAA 144651 GAATGCAGCA GTCCCCATTC ATATTCCTGC TAGACGCGCT GTAAAGTTTA 144701 CTCCAGGGAA AAGAATGAAG CGCTTGATAG AAACTCCGAA TAAGCATTCT 144751 TAATTCTTGT AGTCTTCTTT GTCTCAGTTG TTAGAGTCAG ACCGGTTTTT 144801 TACCGGGCTT GACTCTAATT TTTGTTATTA TTATCGTTTG GTGCAATGCT 144851 TTTCTGATCA AATTGTGCGT GATAATGGGG CTGCAATCCA GGTTACAACA 144901 TTGTATAGAA GTGTCCCAGA ATTCGAACTT TGATTCACAA GTAAAACAGT 144951 TTATCTATGC GTGCCAAGAT AAGACATTAA GGCAGTCTGT ACTCAAGATT 145001 TTCCGCTACC ATCCTTTACT AAAAATTCAT GATATTGCTC GGGCCGTCTA 145051 TCTTTTGATG GCCTTAGAAG AAGGCGAGGA TTTAGGCTTA AGCTTTTTAA 145101 ATGTACAGCA GTACCCTTCA GGTGCTGTAG AACTGTTTTC TTGTGGGGGA 145151 TTTCCTTGGA AAGGATTACC TTATCCTGCA GAACATGCGG AATTTGGCCT 145201 ACTCCTGTTA CAGATCGCAG AGTTTTATGA AGAGAGTCAG GCATACGTCT 145251 CTAAAATGAG TCATTTTCAA CAGGCACTCT TTGATCACCA AGGGAGCGTC 145301 TTTCCCTCTC TCTGGAGCCA GGAGAACTCT CGACTCCTAA AAGAAAAGAC 145351 AACTCTTAGC CAATCGTTTC TCTTCCAATT AGGAATGCAA ATTCACCCAG 145401 AATACAGTCT TGAGGATCCT GCACTAGGGT TCTGGATGCA AAGAACGCGT 145451 TCTTCATCCG CTTTTGTAGC CGCTTCAGGA TGTCAAAGTA GCTTGGGAGC 145501 GTATTCCTCA GGGGATGTCG GTGTTATCGC TTATGGACCT TGCTCTGGAG 145551 ACATTAGTGA TTGTTATTAT TTTGGATGTT GTGGAATCGC TAAAGAGTTC 145601 GTGTGCCAAA AATCTCACCA AACTACAGAG ATTTCTTTTC TCACCTCTAC 145651 AGGAAAGCCT CATCCCAGAA ATACGGGATT TTCCTACCTT CGAGATTCCT 145701 ATGTACATCT GCCGATCCGC TGTAAGATCA CTATTTCCGA CAAGCAATAT

145751 CGCGTGCACG CTGCGTTGGC TGAGGCCACC TCTGCCATGA CGTTTTCTAT TTTCTGTAAG GGGAAGAATT GTCAGGTTGT TGACGGCCCT CGCTTGCGCT 145851 CCTGTTCCCT AGATTCTTAT AAAGGTCCCG GAAACGACAT TATGATTCTT 145901 GGGGAAAATG ACGCAATCAA CATTGTTTCT GCAAGTCCCT ATATGGAAAT 145951 TTTTGCTTTG CAAGGCAAAG AAAAATTTTG GAATGCAGAC TTTTTGATTA 146001 ATATTCCTTA CAAAGAAGAG GGCGTCATGT TAATTTTTGA AAAAAAAGTG 146051 ACCTCTGAGA AAGGAAGATT CTTTACGAAG ATGAATTAAT TTTGGGTCTG 146101 TAATTGTGTT TAAGAATTGT TTGTATTAAA ATGATTCTTT TTATACGAGG 146151 AGAGCACATT CTAATGGAAC TTCTTCCACA CGAAAAACAA GTAGTTGAAT 146201 ATGAAAAGGC TATAGCCGAA TTTAAAGAAA AAAATAAGAA AAATTCTCTC 146251 TTATCTTCTT CAGAGATTCA GAAATTGGAA AAGCGTTTAG ATAAATTAAA 146301 AGAAAGATC TATTCGGATT TGACTCCTTG GGAGCGTGTA CAAATATGTC 146351 GCCACCCTTC GCGTCCCCGT ACTGTCAACT ATATTGAAGG GATGTGTGAG 146401 GAGTTTGTCG AGCTTTGTGG AGATCGCACC TTCCGAGATG ATCCCGCAGT 146451 TGTTGGTGGC TTTGTAAAAA TCCAGGGTCA GCGTTTTGTC CTTATTGGCC 146501 AAGAAAAGGG ATGCGATACA GCGTCACGCC TTCATAGGAA CTTCGGTATG 146551 TTATGTCCCG AGGGTTTCAG AAAAGCCCTT CGCTTAGGAA AACTCGCTGA 146601 AAAGTTTGGC TTGCCTGTGG TCTTTCTTGT CGATACCCCA GGAGCATATC 146651 CTGGATTGAC TGCTGAAGAG AGAGGACAAG GATGGGCAAT TGCCAAAAAT 146701 CTTTTTGAGC TCTCAAGACT TGCCACTCCC GTGATTATTG TCGTTATCGG 146751 TGAGGGATGT TCAGGTGGAG CTTTGGGCAT GGCTGTAGGT GATTCTGTAG 146801 CTATGTTAGA GCATTCCTAT TATTCTGTAA TTTCCCCAGA AGGATGCGCC 146851 TCCATTCTTT GGAAAGATCC TAAGAAAAAT AGCGAAGCAG CTTCCATGTT 146901 GAAAATGCAT GGAGAAAACT TAAAACAATT TGGCATTATC GATACTGTTA 146951 TCAAAGAGCC CATTGGGGGA GCTCACCACG ATCCTGCATT GGTATATAGC 147001 AATGTTCGAG AGTTTATCAT CCAAGAGTGG TTACGATTAA AAGATCTAGC 147051 TATAGAAGAG CTGTTGGAGA AACGGTACGA AAAATTTCGC TCTATAGGTC

147101 TTTATGAAAC TACTTCTGAA AGCGGTCCTG AGGCATAAAA ATCATCTCGT 147151 TATATTAGGC TGTTCTCTAC TCGCAATTTT AGGACTTACC TTTTCATCTC 147201 AGATGGAGAT TTTTTCTTTA GGGATGATTG CTAAAACAGG CCCCGACGCC 147251 TTTTTACTTT TTGGACGTAA GGAATCTGGA AAACTTGTAA AGGTTTCAGA 147301 ACTAAGTCAG AAAGATATTT TAGAGAATTG GCAGGCAATT AGTAAGGATT 147351 CAGAGACACT TACAGTCTCT GATGCCACGA CATACATCGC CGAACATGGG 147401 AAAAGCACAG CCTCTCTGAC GAGCAAGCTC TCTAAGTTTG TCCGTAACTA 147451 CATCGATGTG AGCCGCTTTC GAGGACTGGC AATCTTCTTA ATCTGCGTTG 147501 CTATTTTAA AGCAGTCACC TTATTTTTCC AACGTTTCCT TGGGCAAGTC 147551 GTTGCTATAC GGGTAAGCCG AGACTTACGT CAGGACTACT TTAAGGCCCT 147601 ACAACAACTC CCCATGACCT TCTTCCATGA TCATGATATC GGTAATTTAA 147651 GTAATCGTGT CATGACAGAT TCTGCAAGCA TTGCCTTAGC AGTAAACTCT 147701 TTAATGATTA ACTACATTCA AGCCCCAATT ACCTTCATAT TGACATTGGG 147751 AGTCTGTCTG TCGATTTCAT GGAAGTTTTC AATTCTTATT TGTGTTGCCT 147801 TTCCTATCTT TATCCTTCCC ATTGTCGTGA TCGCTAGAAA GATCAAAAAT 147851 TTAGCAAAAC GTATTCAAAA GAGTCAGGAT TCATTTTCCT CCGTTCTTTA 147901 TGATTTTCTT GCTGGGGTTA TGACAGTAAA AGTCTTTCGT ACAGAAAAAT 147951 TTGCCTTCAC AAAATATTGT GAGCATAACA ATAAGATTTC TGCTTTAGAG 148001 GAGAAAAGTG CTGCTTACGG TTTGCTTCCA CGACCCCTCC TGCATACCAT 148051 AGCTTCTTTA TTTTTTGCTT TTGTCGTCGT TATCGGAATT TATAAATTTG 148101 CTATTCCTCC CGAAGAACTT ATCGTATTTT GTGGTTTGCT CTACCTAATC 148151 TACGACCCTA TTAAGAAGTT CGGGGATGAA AATACCTCCA TCATGAGGGG 148201 ATGTGCTGCT GCGGAGAGAT TTTATGAAGT CTTGAATCAC CCCGATCTTC 148251 ATAGTCAAAA AGAAAGAGAA ATCGAGTTCC TTGGACTTTC TAATACAATC 148301 ACATTCGAGA ATGTTTCCTT CGGCTATCAG GAAGATAAGC ACATCCTCAA 148351 AAATCTAAGC TTTACCTTAC ATAAAGGCGA AGCTCTAGGC ATTGTAGGAC 148401 CTACAGGATC TGGAAAAACA ACACTTGTTA AATTACTTCC TAGGCTCTAC

148451 GAAGTCTCCC AAGGAAAGAT TCTTATCGAC TCTCTTCCTA TTACGGAATA 148501 TAACAAAGGG TCCTTAAGGA ATCACATCGC CTGTGTATTA CAGAATCCTT 148551 TCTTATTCTA TGATACTGTA TGGAATAACC TTACCTGTGG TAAGGATATG 148601 GAGGAGGAGG CTGTTTTAGA AGCTCTAAAA CGTGCCTACG CTGATGAGTT 148651 TATTTTAAAG CTCCCTAAAG GAGTCCATAG CGTGCTCGAA GAATCTGGGA 148701 AGAATCTCTC AGGAGGACAG CAGCAACGTT TGGCAATAGC ACGTGCTCTG 148751 TTGAAAAACG CCTCCATCTT AATTTTAGAT GAGGCAACGT CAGCTCTAGA 148801 TGCCATTAGT GAAAATTACA TTAAGAATAT CATTGGAGAG CTTAAAGGAC 148851 AGTGCACACA AATCATTATT GCCCACAAGC TGACCACTCT TGAACATGTA 148901 GATCGCGTGC TCTACATAGA AAAATGGTCAA AAAATTGCCG AAGGCACAAA 148951 AGAAGAACTC TTACAGACGT GTCCTGAATT TTTAAAAATG TGGGAGCTCT 149001 CAGGGACTAA AGAATATAAC AGGGTCTTTG TTCCTGATCA CAAATTAGTC 149051 GCAAATCCTA CGGACATGGC AATAACAACT TAGGTGGGAT CGCTCTCTCC 149101 ATGAGCTCAG GCAACAACTC TACAAGTGTC TGAGTTAGCT TTTGTGATAC 149151 CTCCTCCAAT CTGCTGAAGG GACAGTCTCC TGGAACAGTA TAATCAGAAG 149201 TGATCTTGAG AAAAGAACAG GGGATGTGAT GTTCTGCTGC TTGTGAGGCT 149251 ATAGCATAGC CTTCCATATC TAGAAGTTTA AACGTCTTAT GAAACCCATA 149301 ATGGTACAAT ACTGGAGAGG TAACCAGAGA GCTTTTAGGT AGAGAATCCG 149351 GTAGAGCGTC AAAGATATAA GGGGGATCTT CAGAGAGAAC AGGAGGTGTA 149401 TCCGTAGTGA GGTTTGCAAT TTTCTCAATA GTGTAACATT GACCTAAAGG 149451 AATCTCGGGA GAACATGCCC CCACAAAACC TGGATTGATC CACAGATCGT 149501 AATCTGTATA TGCTTGGCAA TAGCTTTGAA GAGCATTTAA AACGGCTGTA 149551 CTTCCCCAAA CATGGACAAT ATAGAGATCT AGATGGTAGT CAGTACAACG 149601 ATAACTATAG AGATGCTCGT TGATCTGTGT AAAATCAAGT TGTTCAATTA 149651 GAGGAGAAAT TTCTCTATAG TCTGCAACAA TGCAAAGGAT TTTTTTAGGT 149701 GTATTGACAG CATTCATTGG CCTTCCAGAG CATATGTAAA GCTTTTTTCC 149751 CAGTTTTAGA TAGTTGAAAG GTTTCTTTGT TGATATAGGT TCCTATGAAT

149801 CTATGAATCA CGGTCACGTT TTTATTTTTA GAGTATTCTA CTGCTTTTGC 149851 TCCCGCAGTT ATAGGATCTT TCAGGGAGCA AATTAAAGAC TTTCTTAATG 149901 CTGCTGTTAG AGCATCCACT GTAGCCATAG GAACATATTT CGCAATGGCT 149951 AAACATCCTA AAGGAAGGGG AAAGATGGTC TTACGGCGCC ATAGCTCTCC 150001 AAAGTCTGCC CGCAATGTCA ATTGGAGATC GTAGCTGAAG CGCTCTTCAT 150051 GAATCAGAGC GCCTCCATCG ACTTTCCCTT GCAGTATCGC GGATAGAATT 150101 TTGTCATAAG GCATGGGAAT GAGTTTTGCC TTGGGATAGT AAAGTTTACA 150151 GAGAGCATGA GCGGTTGTCA TCTCTCCAGG AGTTGCCAAG GTATCTAGAG 150201 AACATTCAGG ATCTAAGGAG AGGACGATAG GACCGCTGTT GTATCCTAAG 150251 GTATTTCCTA CGTCCATAAG ATTATAATAA TCAGAAACTA GAGGGAAGAG 150301 CGCTGCTGAC ATTTTCATTA GGGAGAGCCG TCGCTGCAGA GCTAGGGTAT 150351 TCAAAGTTTC AATATCCGCA ATTGTTACCT GGTTAAGAAG AGGCCTGAAT 150401 TGGGGGTCTT TTAAGAAAGA ACGAAAAAGG AAAATATCAT TCGGGCAAGG 150451 AGAAAAGGCA GCAGTCAGTA TCATGTCGGT TGATGTAATA GAGCTATAGC 150501 GGCTTTGATG TCTTTATTTT CAGGCTTATC CAAAGCTCCT TGGTTTTCCA 150551 GCCATTCGAA GTAAGACTTA GGAATATCCA CAAGAGGCTG CCCTTTGTAT 150601 TTGCCAAAAG GCATTTTGAA GACTTTCGGG TGATAGCTCT GTTGCAGCAA 150651 GTCGAGGACT TGCTGGGGCG GTAAATCACC GATTAAAGAA GTAAATACCT 150701 TGTGCAATAT CACTACGTCA TCTAGAGCTC GGTGTGCTTG ATTTTCAGCA 150751 AAACCGTAAA CTTGTCTTAG GTATTGTAAA TTATGTTTTG GTAGATCGGG 150801 GCGATATTTT TGTGCCCATT TTAGAGAGTC TATTGTACGG TTTGTCAGAG 150851 GCTCTAAGGA ATGTCTGCGA CATTCCTTAC CGAGTAGGGG GAAATCAAAA 150901 CCGTCATTAT TATGAGCCAC TAAGATGCTG TCCTCTCCGC AAAATTTCCT 150951 AAATCCCTCG TAGGCTTCAG GAAATTTGGG AGCAGAAAGT ACCGCATCCG 151001 TAGTGATTCC ATGAATTTTG GATGCCTCAT CAGGAATGGG AATTTCCGGA 151051 TTCACATAAG TAAGAAAGGA CTCATCTGTG ACACTATTGT AGGCAGCAAT 151101 TTCTATAATG CGATCTCTTT CTATTTGTGT TCCTGTGGTC TCCGTATCAT

151151 AGAAAATAAG AACATCCATA GTTTGACTAC TCATTACGTT TTTCTTGTTG 151201 CTCTTGAAGA GCCTGACGTC TTAGTTCATC CAAATTCATA TTCCCAGAAG 151251 AGATCAACCC AATAGCATGA GAAAAACTAT CACAGACTAG CTTTATTGTA 151301 TCGATATATA TCCGTAATAG TGTGTCATGA ATTTCTCCGT TTAGGCAGGG 151351 CAACACAGC CGATAAAATA TCAATCCCTG TTCTTCATCC ATGCCAAAGC 151401 CGGGAATATC AATGTCCCTA TTTAAGAGAT GGAGTAAACG AGCTGTTGAT 151451 GCCTTATGAG ATTCATGCAA TTGGTAGGGA AGGTAACAAA TCAACTGCAG 151501 TATTTCTCCC TCACTGCGGA TTACAAAAAA TAAAGGGAGT TCATTGCCAT 151551 TAGCTTGAAT GTTAATGTAA GTAAGACCGC TTTCTCTTTC TAAGAAAGGT 151601 TCTTCATCCG AACTTTTAAG AAATTTTGTG AGATTATTTT GATTTAATGT 151651 CCATGTCGTC ATTTAGGAAA TACTCCAAGT TGTTCCTAGA GCCTGCATCA 151701 TTGCTGGCTG ATAATACTAG ATCTAATCTT GATTCGTCTG TTGTTTTTTT 151751 TGTATCTTTT GTATTGCTTT ACTGAGGAAA TCAGAGGTTG CAGAGGAAGA 151801 AGACTGCTTA TTATTTTCTT GAAGTAGATG ATTGATATCA GTGTCGGGAG 151851 GAAACTCTGT CAGTAGCGTA ATTTTTTCTG GTGTTTCTGC AATCACAAGA 151901 GTTTTATTCA CAACTCGAAT GAGGTAAATA GAAGTTTTCG GCGTTAGGGA 151951 ACGTCGTTCT AGGATTTTGA TTTGAGACGA GCCTCCAAAA CCGTGACTTC 152001 TTGATCTCAC AAACTTTTTA AACGCCCAAA CTCCAAAGCC AAAAATTGTT 152051 AAAAGTAGAA TCAAAGATCC TAGCATTTTA AACATTTCTA ATTTCATGCT 152101 TCCTGGGAAC ATTTCATGTA CAGAAATGGG CTCTTGGATC GTTTCTGCAA 152151 GAGCAAGCTC ATCAGAAAGC TTAAAAAACTA AAGAAAAAAG ATTAAAAAAAC 152201 ATGTGTAAGA CCCGCGATCA TCTCTATAAA ATTATAGTGG TAGCCCGATT 152251 TGTATCCAAC TACATACAAG TAATAATGAA GTATAGTTTT AGTCGATGCT 152301 ATATAAATTA TAGTACAATG ATTTCCAAGT ACAGATTAAA CCGTAATCAT 152351 GTATATTCCT GCAAGTACCG TCTAGAGAGC TCCCCTAGAT GATTTGGGTA 152401 ATTCACAGAC TCCTTTATAC CCTTCTAGGG TGTGCTCGTT CCACAGAGCC 152451 CAAGCTCTTG TCTTTCATAG ACAAAACGAC AGCAGTCTGT CGGTGGATGC

152501 AAAGAATTTT TTCAAGTTGC CTGAGAAATT CCTTGGCAGC GCTTAATAAA 152551 ACTATGGTGA TGCTATGGAA AAGTTACTAG TGACTGATAT TGACGGTACA 152601 ATTACCCATC AATCTCATCA TTTAGATAAA AAGGTGTATG AGCGGCTCTA 152651 TGCGCTGCAC CAAGCTGGTT GGAAGTTGTT TTTCTTGACG GGAAGGTATT 152701 ATAAATATGC TGCACGCTTG TTTTCTGATT TTGATGCTCC ATATTTATTA 152751 GGATGCCAAA ACGGCGCTTC TGTATGGTCT TCAACATCAT CAAATCTTCT 152801 CTATTCTAAA AGTTTACCCT CAGATTTATT ATGTATTTTA CAAGATTGTA 152851 TGGAGGGGC AACGGCTCTT TTTTCCGTGG AATCAGGAGC TCCTTACGGG 152901 GATCACTACT ATCGCTTTTC ACCGACTCCT ATAGCTCAAG ATTTACACGA 152951 ATATGTAGAT CCTAGGTACT TTCCTAATGC TAAGGAAAGA GAGATCCTAT 153001 TTGAAACGCG CTCTTTAAAA GACGACTATG CTTTTCCTAG TTTTGCTGCA 153051 GCAAAAGTCT TTGGACTGCG AGATGAGGTC ATCAGAATTC AAAAGGAGCT 153101 GGAACGCCAA GAAGCACTGA CTTCAGTCGC GACGATGACG TTAATGCGCT 153151 GGCCCTTTGA CTTTCGCTAT GCCATCTTGT TTTTAACAGA TAAAAGCGTC 153201 TCTAAAGGCA AAGCCTTAGA TCGTGTTGTC AATATACTTT ATGATGGAAA 153251 GAAACCCTTT GTCATGGCTT CAGGAGATGA TGCTAATGAT CTCGATCTTA 153301 TTGAGAGAGG AGATTTTAAA ATTGTGATGA GTTCCGCACC TGAAGAGATG 153351 CACGTTCATG CGGACTTTCT AGCTCCCCCA GCAGATAAGA ATGGCATTCT 153401 TTCAGCTTGG GAAGCTGGTG TCCGCTATTA TGACGACCTT ATGAGTCTTT 153451 AGGGAACATC TCAGGACCAA TTCCCATCAC ATTGGCTCCG TGATCTACGT 153501 ATAAGGTCTC ACCAGTAATT GCTGAAGCTA GAGGTGATGC TAAGAAAGCT 153551 GCAACGGCAC CCACCTGCTC GGCATTCATA GCCTCGGGAA TAGGCGCCCA 153601 CTCTTGGTAA TAGTCTACCA TTCTTTCAAT AAAACCAATT GCTTTTCCAG 153651 CTCGGCTTGC TAAAGGTCCT GCAGAGATGG TATTGACACG TATGCCCCAA 153701 CGGCGTCCCG CTTCCCAAGC AAGAGTTTTG GTGTCACTTT CCAAAGCTGC 153751 TTTTGCCGAA CTCATGCCCC CTCCGTATCC AGGAACAGCG CGCATAGAAG 153801 CCAAATAGGT GAGCGATATT GTCGATCCAC CACGGTTCAT GATACTTCCA

153851 AAGTGAGAGA GAAGGCTAAC AAAAGAATAA CTAGAGGCAC TGAGAGCCGC 153901 TAAGTAACCT TTTCTTGATG TTTCTAATAG AGACTTAGAA ATTTCAGGAC 153951 TATTTGCCAG CGAGTGGACA AGAATGTCAA TATGACCAAA ATCTTTTTT 154001 ACCTGTTCTG CGACTTCTGA TATCGTGAAT CCCGTAATGC CCTTGTAACG 154051 TTTATTTCA GCAATATCTT CAGGAACATC TTCAGGGCTA TCAAAACTTG 154101 CGTCCATGGG ATAGATCTTA GCAATCTCTA AGAGAGTGCC ATTCGATAAT 154151 TTTCTAGATT CATTGAATTT TCCTAATTCC CAAGACTGAG AGAAAATTTT 154201 GTAAATCGGT ACCCATGTTC CTACAATAAT CGTAGCTCCT GCTTCTGCAA 154251 GAAGTTTAGC AATACCCCAG CCATATCCTT GGTCATCACC AATGCCCGCA 154301 ACAAATGCTA CCTTTCCTGT TAGATCAATC TTTAGCATGA ATCCGCCTTA 154351 TACTTTTGAA GCTTATTGGA AGGAGAGTAA CAAATCTTTC GATTATTAAG 154401 AAAACCTTTT GGTGCCTCAA CAGGGGAGAT CCTGCCTCCA ATGTAAATAG 154451 AAACGTAAAT TCTTTAAATT TTTTTCTTTA CATATTTTAT AGAATATCCA 154501 AACTTCTCAC TCCCGCGTAC TGCTAAAAAA ATTTTCAAAA GAATTTACGA 154551 TCCGAACTTA TCGTAGTTTG GGTTTCACTG ATTACTTAGG AGGTTGTTTG 154601 ACGAATCCTT TAGGGAAATT CCCCTCACCA CAGAATCCAC AGGTTGTTAC 154651 GATAGCGCCT TCTTCCACAA CACCACAAGC AGTCTCATCT GCAGTTCAAG 154701 GTTTTCTTCA AACTGGAGGA GCTGCCTCCT CTACAGCGAC AACTACTACC 154751 GCATCCGGAG CCTCTGCATT AGGACTTTCA CCTGATCAAG TGCAAGCGTT 154801 GCTTACTAAT TTATTAAATG TGGGACAACC ATCAGTGGGA CAACCATCAA 154851 CTTCAGCAGG AACTTCGGGA GCCTCCTCTT CCAGTGCAAG TATGCAGCAA 154901 CAGCTTTTGC AACTTATCTT AGACAAGACA ACAGGAAGTG GCGGATCGTC 154951 CGTGAGTTCA GAGCAATTAC AGCAACTCCT TAGCTTGGTG AGCCAGATGA 155001 CTACGTCTCA AGGAGGAAGT GGTGGAACTC AGGCAGGACA GGCCGCTTCG 155051 GTACTGTTGA ATTTGTTATC GGCAACAGGA TCTGCAGCAG CAAATCCTTT 155101 AGGGACAGCT GCATCGTTGG CACAGATCAT TTATGCAGCA GTAACAAGTC 155151 CTGGAGCAAA GAAAACTAGC GAATTTTGTT ATAATTATTG TGGAGAGACC

155201 TGCCAAGGCA ACTGCGGTTG TCCTACCTGT GGCTGTCCAG ACGGACAGTG 155251 CGGTTGTGGA GGATTTGGCC GTTTTTTCTG TGGTGTATGG AAAAATTGTT 155301 GCGGGATAGG AGAGGGATCC CAAGAACCCG CAATCCCTTT ATAAGAACTC 155351 GAGGCTTTAG AACAGAAATA TGGCAAGGCT GTTCTTTAA TTGCGTTAAG 155401 TGAGCTTGGC ATTGATACCA TGAGCTTATT ATCAGGACAT CGACTCGAGG 155451 GATTTCCTCC AATCGCGGAG GTCATGGCTG CATGTGACCG GTGTTCTATG 155501 GACTTTTGTG AGATCTTGAA GTCTCAAAGC ATGGATCTGT GGGCGGATGC 155551 GGCGAGTTGT GTGGATGGTT TATTACAGGA TCCTTTTTGG AGTACAGCAA 155601 TTGCCTCAGG GATTGCTAAG TCTTCTCTTC AGGAAACGGA ATTCGAGTGT 155651 GAAAGCAAAG TGATGGTTCT TTCTTCATGG GGAGAGCAAG GAGCACAGGT 155701 TTGTAGCCCT TTTAACCTAG AGAGGATATG TATGTCTTTC CCATCACTTA 155751 AGGTCTTCTC CCTTAAAAAG AACGGGTGCG AGAACATGGG AATCCAGTTG 155801 TCTGCATCCT GCATGAATCT ATTAATGTCT ATTTTCTTTG TAGCTACCAA 155851 TGGAGGAAGC ACTCCGATTT GGATCACCAA AGAAAATCTG ATGGCGTTAG 155901 TTGCTTTGGT TTTATCTCAC TATCAATGTT ATTTTGTCCC AGCCACAGGA 155951 GATCCCCAAC GTGGCAACAT TTTAGGTAAT CCAGAAGTCA ATGCTATTTT 156001 GGCTCGGGGG ATGGGCATGC GTGTCGATCT GGAAAGGAAG CGAGGGGGAG 156051 AATCTTCCTC GTCACGCTAT TTAGAATTAG CTGCACGATG TTTTGAGAAT 156101 TCTCTTACGA AAACAAGTTT GTTAAGCGAT GCTAACAATG TTCAAGAAAG 156151 AGATAAGTGC CTACTACAGA TGTCAACTTC ATTGATGCAT ACGGCGGGAC 156201 TAAATTTACA ACGCCCCCCT GTACCCACAC CTTCTGGAGT CACGGCACAT 156251 CCGCAACCTC AACCAGATCC TGTGGTTACG TCTCAACCTT CTTTATTAGG 156301 TGCTAGAGAG 'CGTTCCCCTG TGTCTTCTAG AGGGCGTTTT CCTGTAGTTT 156351 TACCTTTAAG TGTGATTTCT CCTAGGTCGC ACCCCGGAAG GGTAGAAAGG 156401 CGGGATTTAG AAGATGAAGA AGAGGAGGTT ATGTTTTGAA GCAGTGTAAA 156451 CGACTCCAAT TACAGTTTTA TGAATCTCTA ATTGTAAAGT TCTAGGGGTT 156501 TTTCTTGAAG TAAGTGCCGA GCACATTCTC TAGGATCTTC GGTTGATGAC

156551 GCACAAATTT TTAGGGGAAG ATTTGTGAAT GGGGAGATAA ATTCTAGGGA 156601 GTGAGCATGG AGGAGAGGGC GGAAGATCTG GGGAGGCTGT TCTTTAGGTC 156651 CGTAGTCGAC ATCTCCGACA ATAGGATGAC CCAGCAATCC CATTTGTAAG 156701 CGGATTTGAT GGGTTCTCCC TGTGATGGGC CTGCGGCTCC AAAAATCACA 156751 GCTCCACACC TCCGGTATAC GGGGGCCGTA TAAGATTTTA CGGTTCCAAA 156801 TTTTTTTTA GGATGACCAA AAACGAAAGC TATGTATTGT TTATGGATTT 156851 TTCTTTGCTT GAACAATTTC ATGAGCTCAG TAGCCGCTTG TTTAGACTTT 156901 CCCATGAGAA GACACCCAGA GGTGCCTTTG TCTAACCTAT GCACAGTAAA 156951 AAACCGTGTC ATGTGTGCCA TTTGTTCAGT AGTAAGATGG GGAGGTTTTT 157001 CGTAGATAAT GCTATAGTCA TCCTCCCAGA GGATGCTAGG TTGTTGTTTT 157051 GTTGAGGGGA TCAGAGATAG GGAAACACGG TCGCCAGGTT GTACCTTGTA 157101 GGATTCAAAT CTTTCTATGA ACCCGTTCAC TCGACATCGA TGTTGGCGAA 157151 TAGACGCCAA GATTTCTTGC TTGCTATGAT TAGGCAGTTG AGATCTAAGA 157201 AAAGAAGATA ATCTTGAGAC TTGTGTGGCA AGCCAGGAAA AATTTTCCAT 157251 AAAATATTGT AAAGCAGCCC TTTTATCATT TGATAATTGC ATAAAATTTT 157301 AAGAGATTTT GTATGACAAA GATAGCTTTT TCTGAAAAGG CAAAGAATTT 157351 TCCTGTAGAG GCATTAAAAA AATGGTTTGA AAAAAATAAA CGATCTCTTC 157401 CTTGGAGAGA TAACCCGACT CCCTATAGTG TGTGGGTTTC CGAAGTTATG 157451 CTACAGCAAA CGCGAGCTGA AGTTGTTATA GATTATTTTA ATCAGTGGAT 157501 GGAGAGATTT CCTACCATAG AGTCTTTAGC TGCAGCAAAA GAAGAAGATG 157551 TCATTAAGTT ATGGGAGGGA TTGGGTTATT ATTCTCGAGC GCGCCATCTT 157601 TTAGAGGGAG CTCGCATGGT TATGGAGGAG TTTCATGGAA AGATCCCTGA 157651 TGATGCCATT TCCTTAGCTC AAATTCGTGG AGTTGGTCCT TATACGGTTC 157701 ATGCTATTCT AGCCTTTGCT TTTAAGAGGC GTGCTGCTGC TGTGGATGGC 157751 AATGTCTTGC GTGTTCTTAG CCGGATATTT TTGATAGAAA CTTCTATAGA 157801 CTTAGAATCA ACTCGTACTT GGGTTTCTAG GATTGCTCAA GCGCTTCTTC 157851 CTCATAAGAG TCCCGAGGTT ATAGCTGAGG CTCTGATAGA GTTGGGAGCT

157901 TGTATCTGTA AAAAAGTTCC TCAATGTCAT CGTTGTCCTG TCCGTCAAGC 157951 ATGTGGAGCT TGGAGGGAGA ACAAACAGTT CGTATTGCCG GTACGTCATG 158001 CCAGAAAAA GGTCATCTTT TTGCATCGTT TGGTAGCGAT TGTATTGTAC 158051 GATGGCTCTT TGGTTGTCGA GAAGAGACGT CCTAAAGAAA TGATGGCAGG 158101 CTTATATGAA TTTCCTTATA TTGAAGTTGA ACCAGAGGAA GGTCTTCAAG 158151 ATATAGAAGG ATTTACTAAG AAGATGGAGC TTTCTTTAGA AAGCCCTTTG 158201 GAATTCTTAG GTAACCTTAA AGAACAGCGG CATGCGTTTA CTAATCATAA 158251 GGTTCATTTG TGTCCTATAA TTTTTAAAGC CACTTCTCTG CCTCAGTTCG 158301 GGGAATTGCA TCTTTTGAGT GATATAGATC ACTTAGCTTT TTCTTCAGGA 158351 CACAAAAGA TTAAAGATGC TTTGCTAATC TACCTCGGGG ATGTCAGGTC 158401 TAGAGAATCA ATAGGAGTAT AGATGCGAGA TCACGCTTTT TCTAAATTGA 158451 TAGGGACTGT CCGTGCCATG GTAGTTGAAG GACGTTGTCC TTGGTCACTT 158501 CAGCAATCCC TAGTCTCTAT GGTAGAGCAT ATTCTTGGAG AGTGTCAGGA 158551 ATTTCACGAG GCCGTCTTAC AAGGTAAGAC GGTACAAGAG GTTGGTTCCG 158601 AAGCCGGGGA TGTCTTAACT TTAGTTCTAA TTTTATGTTT TCTGTTAGAA 158651 CGAGAGGCC TACTTGCTTC CGAAGACGTT GCCAATGAGG CTATGGAAAA 158701 ATTGCGTCGC CGTGCTCCTT ATATATTCGC TGAAGATTAC AAGCCGGTCT 158751 CGATTGAAGA GGCCGATCGC CTTTGGGAGC TTGCTAAGCA CCGAGAGAAA 158801 AATGAATCTA CATAGTTGAA GTTTTGGTCT ATTTTTAAGC ATATGGTGCT 158851 TTTGAAAAAA CAGAATATAT GCTATCAAAG AAGGGTAAGT TGGGGGCCTT 158901 TTAAGAGAAG GAACCTGCGA ATCGGGTCAG GACTGGAAGG TAGCAGCCCT 158951 AAGGAGAGTT TTCTTTTGCT AAAAGAATGT TCTCCAACTT ACTCTTTTTA 159001 CTTTATTCCC AAAAATAGCA ATGAGGTGAG GTTAAACAAC CCGTGCAGTG 159051 CAATGGGAGA AAGAATGTGC CGATCTTTTT CATATAGAAA CCCTGCAGAT 159101 AAGGAAAAA CAAAGAGCAC GGGGACAAAG ACCCAACTTC CTAAAGAGTG 159151 TTCAATGTGA ATGAAAGAGA AAATAATAGA AGAGCATAGT ACCGCAGCTA 159201 TGCGCGTCAT TTTGTTTTTC AAGAATGTCT GTAGAATTCC TCTAAAAAAT

159251 ACCTCTTCTC CAAATGGAGT GAGGACGCCT AAATTTAGAA TCATGCTAAT 159301 GTAGTGTCCT GTTATAGGCA GAGAGTTCTG AACTTCTTGA GTGACTTCTT 159351 GTGTGTGAAT CTCTTGCGTA GGAAGAACCA AAGTTAAAAA TTTACTCATC 159401 ATAATCCCAA TCAGTTGTGT TACTGGGATG ATGATGATCC ACATTCTGAT 159451 GGCAGATCCT AGAGCACGCC ATGAAGTTTT AACCGGTCTT TCTCCAGAGA 159501 AAAGTATAGC ACGTGTGATA TCCTTGGGGA GAAAAAGCAG GTAGAACAGA 159551 AATGCAAAGG CAAGGCTAAT TCCTGTCATG GTGGAAAGTA ATTCCGCAGT 159601 TTGTGAGCTC ACGCTAAGAG CTACAAGGGA AGAAAAACA AGAAGAGCAC 159651 CACCAAATAA AACTTGGCGG AGCTTTAAAG GTGTTTTCCC AGAGGGTGCT 159701 GGCCAGATAA AGAAGTTTTT GGAAGCTAGA GCAGCGACGC CAAGGGACAA 159751 GAGAAGAATA AACTTGGACA TTTCCTTAGA CTACGAGTAG TTAGCACAAA 159801 CATAGCCCTC AACTCTGGCA ACAACTTCGC GGAAAAGACG GCTATGCATT 159851 AAGCCCATGG GCGTTGAATC AAAATGTTTT GAGTTCCAGC CATAGCGATG 159901 ATAATCATTG ACTAGGGTAG TTAAAGGCTG GCTGCATTCG ATAATCTCTT 159951 GATAAATGAG AGCTATTTTA TGATGACGGA TATCAAAAAC GCGAACACGT 160001 ACAGACGCTG TTACAGAATC GACACCTGCT TCTTTCCCTG TCTTTTGTTC 160051 TAACAGTTCT GTAGCAACAA TGAATTCTGC AGGAAGAAAT TGCTCAATAA 160101 TTGTTTCGGG TAGACGATTC GCAATCGGAG CATAGAACTG AGAGACTGTC 160151 TGAGGTGAAG CATTGTGCTT GATCAGGAAG ACCTTTTCCG AAGCATAAAA 160201 CCTTTTGCTG ATCTCTTCAG TAAATTCTCC TTGGAGGTTC CAAGGTAAAG 160251 GTTCAAGACT CTTTCCTGGG CGATGAAATA CAGGAAGCAT CGCAATCACA 160301 CCTTTAGTTT TGCTCCCTGA AGTGTATAGC TTAGGATGAT AACTTCCTGA 160351 AGAGCCTAAG TGAGTGCAGC TGGATAGGGT TGGGGATAGA AGTCCTAAAG 160401 ATGCCAATAA TACCAACATT TTTCGCATAG TCACTGTCCT TAAATTGCTT 160451 ATTTTGCAAA AGATTCTAGC CCTGGGAAAG TTTTTACTTT TAAGATCAAT 160501 ACTTTCGCAA TTGAGAGATT TTCCATTTAA AACTCTCATT AGCTTATATC 160551 AAAGAAAAA ATAAAAACAA GCAAAGAGAC CGTCTCAGTT TTAGTTTAGA

160601 AACTCAAGGT TGAGAAAGGG ATTCTGACCA AAGTTGTGAG GGAACTTTGG 160651 TAACTTTTC TTTAGGAATC AATGTGCACC CTGGGAGGAA TACAACTGGA 160701 GCTGAGATCA CAGAAAAAT AAACCACTTC ATGACTACCT CTGCAAAAAG 160751 AATACTACTA TTTTCTATTT GCATAGGCAG TTCTTCGATT TATAGCAATT 160801 TTTACTTTAT CATCATAAAA ACTATGATGA AAAGGTTCTT AGTGAACTTC 160851 TAAGGAACAC CATGAGTTTG TGATTAAAAC TCATGGTGCG TAGTTGAACC 160901 CTTATTGAGG AGGGACGCAA CCACTAAGAG CTAACAATAC TAAACAAGAT 160951 AATAGTAAAG AGAATGCTTT TTTCATTATT CATCCTTAGG TTAATTGACC 161001 TTTCCAGCCT AGCTCTAGGG CGATTATTTA TCAAATTTTT CTTTGTAATT 161051 AATGATCATG CGACCATTAA TTTAGCGATA AATTATGATT TCGTCAGGAA 161101 AATTCAATTC TTTATAATAA TGATATGAAA TTAGAGAATG TCTATAGGGG 161151 CGGACTCTAT TTGTGATCCA GGATCTCTTT AGGAGCACTT TGTGGATTTT 161201 GATTATTTTG GTCTGAGTGA TATTGGTAGG GTGCGCGCTA GAAATGAAGA 161251 TTTTTGGCAG GTAAACCTCA TGTCTCAAGT GGTTGCTATT GCTGACGGTG 161301 TTGGGGGGCG TCTTGGTGGA GACATTGCTT CTCAAGAGGC AGTGACTAGC 161351 CTTATGGAGC TGATTGATGA GCAACAGTCA AAATTGATGG GGTATGGGGA 161401 TGACCAGTAT AAGGAGACTT TAAAAAAGAT CCTTTTAGAG GTCAATGGTG 161451 TGGTCTATGA ACACGGCCAA ATGGAAGAGC ATCTCCAGGG TATGGGAACC 161501 ACTCTTAGCT TCATCCAATT CCGGAAGGAT AGGGCATGGC TATTTCATGT 161551 GGGAGATAGT CGAATTTATC GTATTCGTGA GGGAGAACTG CGCCGCCTTA 161601 CCGAAGACCA TTCTTTAGAA AATCAATTAA AAAATCGTTA TGGGCTTCCT 161651 AAACAATCAG ATAAGGTGTA TTCTTATCGC CATATTCTGA CTAATGTTTT 161701 GGGAAGTCGT CCCTATGTCA TGCCTGACAT TCGGAATCTT CCTTGTGAAA 161751 AGGAAGATTT GTACTGCCTC TGTTCGGATG GATTGACAAA CATGGTTCCA 161801 GATATCGATA TTCGTGATAT CTTGAACCAG CCCGCCACCC TAGAAGAACG 161851 GGGGAATGCA TTAATTTCTC TAGCCAATAC TCGTGGAGGC GATGACAACG 161901 CTACTGTCGT ATTAGTCCGA ATACAATAGT TCCTTTGCTA AGGATAGTAT

161951 TCCATGATCT ATTTGGATAA CAATGCGATG ACACCCCCAG AGAGGGGACT 162001 TTTGGAATTT CTCCAAAAAA CCTTCCTTAT AGAAGGGACG TACGCGAATC 162051 CTTCGAGCGT CCATCAATTA GGTAAAAAAT CTCGTCAACT GGTTCTAGAA 162101 GCTTCACACT GGATGCAAAA GGTCCTTTCG TTTCAGGGCC GTGTCCTCTA 162151 TACCTCAGGG GCTACTGAGA GTTTAAATTT AGCAATAGCA AGCCTCCCTA 162201 AAGACAGTCA TGTTATCACC TCAGGTAGCG AACACCCCGC CATCTTAGAG 162251 CCTTTAAAAC ATTCCTCGCT TTCCGTTTCT TATTTAAATC CCGAAGAAGG 162301 GAGATGTGTT CTTACTATAG AGCAGATTGA AAGAGCTGTG ACTCCTAAAA 162351 CTTCAGCAAT CATCTTAGGT TGGGTCAATA GTGAGACTGG TGCCAAAGCT 162401 GATATAGCTG CTATAGCCCA CTTCGCGCAA GAACGACAAT TGCAATTTAT 162451 TGTGGATGCG ACTGCAAATG TAGGTAAGGA GAGGATAGTT CTTCCCTCTG 162501 GTGTCACTAT GGCAGCATTC AGTGGACATA AATTTCATGC ACTCTCTGGA 162551 ATCGGAGCTC TTCTGGTCTC TCCAGGAGTC AAACTACATC CTCAGCTGTG 162601 GGGAGGAGGT CAGCAAGGAG GGCTGCGCGC AGGCACAGAA AATCTTTGGG 162651 GAATCGCCTC TCTGCTTTAT ATTTTCAAAT ACCTAGATCT TCATCAAGAG 162701 CGTATCTCTC AGGAAATTCT TACCCATAGA AATGGTTTTG AAAAGGCAAT 162751 CAAAGCACGC ATTCCTGATG TCCATATTCA TTGTGCGGAT CAACCACGGG 162801 CAAACAACGT CTCAGCAATT GCTTTCCCTC CGTTGGAAGG TGAGGTATTG 162851 CAAATCGCCT TAGATATAGA AGGAGTGGCT TGTGGTTATG GATCCGCATG 162901 CTCTTCAGGT GCTACCGCAC CCTTTAAATC TCTTGTCAGC ATGGGTGTTG 162951 ATGAAGAGTT GACCCTGGCA ACACTCAGGT TTTCTTTTAG CCATCTTCTC 163001 TTGCAAGAAG ATGTTGAAAG AGCCGTTGGA ATTATAGAAA AAGTCGTAGA 163051 ACGTTTGAAA AATTCCTAAG TCTTAAAAGA GAACATGTTT CTAAGCTGAA 163101 AGAACACTCC TGACTCTTAT TGCAGAATCT ATGAGAGTAA GTTTTTAATC 163151 GATACGGTTT TTATCCCAGA TAAAGACATC TCTTTAACTT CTAAAAGCAA 163201 GTTGTTGATA ATTACAGAAG TTCCTACTTC TGCAGGACTG TCTAGCAGTT 163251 GCAATACCAA TTGGGCTAGG GTTTCTACAG GATATTGCGG AAATTGAATA

163301 TCGAGTTCTT TTTGCAGATC TTTTATGCGA GAGTTGCCAG GAAACGTTCT 163351 TTCAATAACA GAGATGGTCT TGGGTTTTAA ATGAGCAATG TTTGTAGTGT 163401 TGAATAAGAT TTTGAAAATT GCATTTAAAC TAAGAATACC TATAGGTTCA 163451 CCAGAAGCAT TGAGGACAAC AGCAACACTC GAACGGTTGT CTCGAAACTC 163501 TTTGAGGATA CGAATAAGTT TTGATTTTGC AGTGATAAAC CAAGGCGAGT 163551 GTAGATTATT GATTAGGGGT TCATCAAGAG CTTTATTGAC AAAGTCTTTA 163601 GGATGGGCAA TCCCAATAAC GTTTTTTCGG GCCTTGTGAT AGACAGGAAT 163651 AAAGTTGATA TCTGTATTTT TTATAGTCCG GCAAAAATCT TTAACATTTG 163701 CAGAAGAAGG AAGCATGGTA ACCTGTTCTA AAGGTTGGCA TACCTGATCT 163751 GCACAAGTCG CACTTAAAGA GAAAATATTT GTAGCAATTG TATTGAAATC 163801 TTGTTCTTCA TGGTGAGTCT CTAAAGCTTT TTGGAACTCG TCTCTACTTA 163851 ATGTAGAGTT CAATTTTCT TTCCTAATAT TTAGAAGATA GTAAAGACCC 163901 TCAGTGAGAC TTCCTATGAG CTGAATCAGA GGATAGAAAA TATAGTGGGA 163951 ATAATAGAGA ATCGGTGCTC CCCAAAGTGC TAATTTTTCA GGAATCTTCC 164001 GTGATATTGT TAGAGGTAGA AGTTCTGCAA AAATCACAAC TATAAAAATT 164051 TGAGTGAAAG GAGCGTAATC TGGAGTGATT CCTAAAGCTC GATAGCAATT 164101 TCTTGAGGAC TCAGACCCGA CTTGTAGAGC GATATTCACT CCTAACATCA 164151 CCGTTCCAAA TAAACGATAG GGGCGGCGAA TCAGGAAATT AATGTAGCGA 164201 GCTTTCTTAT GATCTTTAGT CAGATAGTAT TGCAATCGTA CACGGTTAAA 164251 TGACACGCAG GCCATTTCCA TCATCGAATA GAATCCTTGT AAGACAATAC 164301 AGATAATGTT GACTCCTATC CAAAAGAGAG CAGAATTAGT CATACAATTT 164351 CCTTATATAC ACACGGCGAA TGCGATTCGG AGCAGCGTCT AATACCTGGA 164401 AAAGCAAGTT ATTCCAAGAG AGTTTCATTC CTGTTGTCGG AATCGTTCCG 164451 ATTTGCTCTA TTAACCAGCC TCCTATAGTC GCAATATTAT TGTTCGTCGG 164501 TAGGTTGATA TCGAAGATCT CACTAAACTC ACGGAGTTCT AAAGTTCCTG 164551 AGGCAATAAT AACATCAGCT CCTGAGGTGG TATAGAGTAT TTTATTATCT 164601 CTCTGGTCTA CAATTTCTCC AGCAACAATT TCAAAGAGGT CTTCTTGAGT

164651 GATCAATCCT TCAATAGATC CGTATTCATC AATGATCATC CCTAGGGTTT 164701 CGTCTTCAGC TGCCATCTGA CATAAAGCCA TTTTTGCAGA GATGGTTTCT 164751 GGCATATAAT ACGGTTTTTT CAGCAAGGGG AGGAGATCAT CCGAAGATTG 164801 CAGTGGCTTG TCATGTAAAA GAAGAGAGCG CGCTGTGCAA ATGCCCAGAA 164851 GGTTTTGGAG GTTATCGTTA CATATAGGAA CTCGTGAGCA ATGCTGTTTA 164901 GAAAATAAAA GATAGAGGTT CTCTAAAGGG GTTTGGATAT CATAAAATAA 164951 AATATCCTGG CGTGGCTGCA TACGCTCTTT AACACTACAA TCACTAAGAG 165001 AAAGATAACC ATAGAGTAAA CGGCTTTCTT CTTGATTGAC TACGCCGAAA 165051 TCCTTACAAC TTTGCAATAC TTCCTTCAGC TCTTGGGGTT GGATGATATC 165101 AATCTGTTGC TTCGATAAAA TCCATTGGAC CACATAATTA ATTCCTACGA 165151 TACCCCAGTG GAGTAGGGGT TTGAAGATTT TAGTAACACA AAGAATAAGA 165201 GGGGCTACGG AACTAGCAAT CTGTGTATTA AAAGGAAGAG CTACTGCTTT 165251 AGGGAGAATC TCACCTAAGA TCAAAGTAAT TGCTAAAGGA AGACCTACAG 165301 TAAACCACCA CGAAGCTGCA TCTCCAAATA GAATGGCAAA ACAGTTTTGA 165351 ATAGCAATAT TCAGTCCGAT ATCACAAAAA ATTAAGGTGA TGAGCAGGTG 165401 GTGGGGATGT AGAAGAAGGG TAGCTACTCG CTGCTGTTTC TTAGATTTAG 165451 AGCGCTTATA GTGCGAGATC AAACTCGTAG GCAAAGAAAA CAAAGCAATT 165501 TGAGATAACG AAATGAATCC CGAGCATAAA GTAAAACAGA TAATGAAGAA 165551 CATTAACATG GTAGGAATCA TGGTCTCTTT TCAGTCCTTA TTTTCTGATT 165601 GTTGCTTTGG GGAGACACAG AGTTTCTTAT AGCCTTTAGG AATGAGACCC 165651 AGACGCTGGA TTAAAGCAGC TTCTATAGCA GCGGAGTCTT GCCAGTGTTG 165701 CAGATGTAAT TGGAGTTGAC GCTGCTTTTC TTGAGCAGAA AGAATGTCTT 165751 GGCATAAAGA AGAGACCTTG CTTTGTAAGC GTAGCTCTTC TGTACGTAAC 165801 TCCTGGATAG CACGATCATA AACAAAGCCT CCAATTAAGA TGCTAAAGAT 165851 CACCCACCAG GATTTGATCA TCACTTCTTC TAGTAATCTA AAACCCCAGT 165901 TTTTTTTCT TACTGAAACT TTAGACACAA GGTACGGTGA TGCCTTGTTG 165951 CTTTTGATAC TTTCCTTTTC TGTCTGCATA AGAAACCTCG CAGGTCGTAC

166001 TAGACTCAAA GAATAAGACC TGGGCAATCC CTTCATTAGC GTAAATTTTC 166051 GCTGGCAATG GCGTAGTGTT AGAAATTTCT ATAGTCACAT GCCCTTCCCA 166101 TTCAGGCTCA AAAGGTGTGA CATTTACGAT AATTCCACAG CGTGCATATG 166151 TAGACTTTCC TATACACATT GTTAAGACAT TTCTAGGAAT TCGGAAATAC 166201 TCAACGCTAC GAGCTAGAGC AAAAGAATTT GGAGGAACAA TACAGACGTC 166251 ATCAGTAATA GAGATGAAGA TATCCTCAGT AAAGCATTTT GGATCAACAA 166301 CAGAGTTATA GACATTGGTG AACACTTTGA ATTCTCGAGA TAGGCGGAGG 166351 TCGTAACCAT AACTCGATAG GCCGTAACTT ATAAGTTTTT CGCCTGTCTC 166401 CTCATTTACG TTCACTTGGC CATTAACAAA GGGATGGATC ATATCGGCAT 166451 TTAGGGCCAT CTCTCGTATC CACTTATCTT CTTTTATGCT CATTTAGAAA 166501 CCTTAACAGT TTGAAATTGC TTTCTTAATG ATATTCTGTT TTTCAATTTA 166551 CTGGTTTTTG GGGGGAACTT TTCTAAGTAT AAGATAGACT TTGATTATCT 166601 CTTGAAAAGA CCAGTTGTAT AAACAAGAAA AGCCTATCCC AAAGGCTACA 166651 ATTTTATCAC GAAACCTAGA GGTAATGTTA GATAATCCTA AGGGAAAAAG 166701 GCAAACCTTA TTTTTAGGGA GAACTTCAGG TAGGTCTGCT CTTTACTCTT 166751 ATAGTAGAAG AATCTTGGTT CTCTTGAATG CATTCATGCG AGGACCTTGA 166801 TAAGAACTTC TTGGATTCAT AAAAAGATTA ACATCTCCTT ATTGATAAGC 166851 TAGAGAATTT TTACTACCAA CTTCTCAGTG GAAAATGTTT TTAAAAATAG 166901 TTCGCCATCT TTAATTTATC TGTTTTAAGA CAAAAGAAAT CTAGATCACC 166951 ACAGGAAGTT TAAATCATAA AATGAAAATG ATGGAGAGGT TCTAGTGCTC 167001 GTACTTTGGC CCTGCTCTCC TTGATAGAAA GAAGAGGTCC ATAGTGTACT 167051 TCTATATAGT ATCTCGTGTA CTATGCCGAG TATAACCGAT CGGCGTTATC 167101 GATGAGAGTT TCAAAAAAAT ATAAAATCCA CCTAAAGAAA AAGCGATAGA 167151 GAAGGTTCGT ACATGACGCA TCAAGTAGCT GTCTTGCATC AGGATAAAAA 167201 ATTTGATGTT TCGTTAAGAC CTAAAGGGTT AGAAGAATTT TATGGACAGC 167251 ATCATTTAAA AGAACGCCTA GATCTATTTC TTTGCGCAGC ATTGCAACGA 167301 GGAGAAGTTC CAGGACATTG CTTGTTTTTT GGACCCCCAG GCTTAGGGAA

167351 AACCTCACTT GCTCACATCG TTGCCTACAC CGTGGGGAAA GGGCTGGTCT 167401 TGGCATCAGG GCCTCAGTTA ATCAAACCCT CGGACCTGTT AGGACTTTTA 167451 ACTAGTTTGC AAGAAGGGGA CGTGTTTTTC ATCGATGAGA TCCATCGTAT 167501 GGGGAAAGTT GCTGAGGAAT ACCTGTATTC TGCAATGGAA GATTTCAAAG 167551 TCGATATTAC TATAGATTCA GGACCCGGAG CTCGCTCGGT CCGTGTCGAT 167601 CTTGCTCCTT TCACTTTAGT GGGGGCAACG ACTCGATCAG GAATGCTAAG 167651 CGAACCTTTA AGAGCACGCT TTGCTTTTAG TGCGAGACTT TCCTATTACT 167701 CGGATCAAGA TCTAAAAGAG ATTTTAGTCC GCTCCTCACA TTTACTCGGA 167751 ATCGAAGCTG ACAGCTCCGC ATTACTAGAA ATTGCTAAGA GATCCCGAGG 167801 GACGCCACGA CTGGCAAATC ATCTTCTACG TTGGGTCAGA GATTTTGCTC 167851 AGATCCGAGA AGGAAACTGT ATCAATGGGG ACGTAGCAGA AAAAGCTTTG 167901 GCTATGCTAT TAATAGATGA TTGGGGATTG AATGAAATTG ATATCAAACT 167951 TCTCACTACA ATCATCGACT ACTACCAAGG TGGTCCCGTT GGAATTAAAA 168001 CCTTATCGGT AGCTGTGGGA GAAGATATCA AAACTCTTGA AGATGTTTAT 168051 GAACCGTTTT TAATTTTAAA AGGTTTTATC AAAAAGACTC CCAGAGGCAG 168101 AATGGTAACA CAACTTGCTT ACGACCATTT AAAAAGACAT GCAAAGAACT 168151 TATTGAGTTT AGGAGAAGGA CAGTGAAACT ATTGAAAAAC GTACTTTTAG 168201 GTCTTTCTT CAGTATGAGT ATCTCAGGAT TCTCAGAAGT AAAGGTATCC 168251 GATACTTTTG TGAAGCAGGA TACTGTCGTT GAACCTAAAA TTCGTGTCCT 168301 TTTATCTAAT GAAAGCACCA CAGCTCTCAT AGAAGCCAAA GGTCCTTATC 168351 GCATTTATGG AGATAATGTC TTATTAGACA CAGCGATTCA AGGCCAGCGT 168401 TGCGTGGTCC ACGCTCTATA CGAAGGGATC CGTTGGGGAG AATTTTATCC 168451 CGGACTCCAG TGTTTAAAGA TCGAGCCTGT AGATGACACT GCTTCTCTTT 168501 TTTTTAACGG GATTCAGTAT CAAGGTTCCC TATACGTTCA TCGTAAAGAC 168551 AACCATTGCA TCATGGTTTC TAACGAAGTT ACAATCGAAG ATTATCTGAA 168601 ATCTGTACTT TCTATAAAGT ACCTTGAAGA GCTAGATAAA GAAGCTCTAT 168651 CTGCTTGCAT CATTCTAGAA AGAACCGCTC TATACGAAAA GCTCCTTGCA

168701 AGAAATCCTC AAAACTTTTG GCATGTTAAA GCTGAAGAAG AAGGGTATGC 168751 AGGATTTGGT GTGACCAAGC AGTTCTATGG TGTAGAAGAG GCTATAGACT 168801 GGACAGCTCG TTTAGTTGTG GATAGCCCTC AAGGATTAAT TATAGATGCA 168851 CAAGGGCTCT TGCAGTCCAA CGTAGATCGT CTTGCTATAG AAGGATTCAA 168901 TGCACGTCAG ATTCTTGAGA AGTTCTACAA GGATGTGGAT TTTGTAGTTA 168951 TAGAATCCTG GAATGAAGAA CTGGACGGAG AGATCAGGTA ACCTCTTTCG 169001 CATGGCTGAT CGCAATTAGC GTGGTATGGG GCTGTAGCGA CACTGTCGGC 169051 GTTGCTACAT TTTGAGGGAC AAACCCTTGC TGACTCTCGG CAACTATTTG 169101 ATAAGGAAGA AAGTTGCTGG AGGCTTTAGG TAAGGTCGCA AGTTGGTCTT 169151 GAGCTCCCAC GTGAAAAGCA ACATATACAT GCGCTTTTGG CGATTTTATT 169201 TTAAATGCTA AGAAATTTCC AGGGCGCCAT GTCATGGGAT TTCCCATAGC 169251 ATCTACCCAA CTGATTTCCT TATTGGAAAG AAAGCCTCGA TTAAAAAGTG 169301 TTTTATATTT TTTTCGAAAC GCAATGAGAT CACAGAGAAA GTGCATCAGT 169351 GTAGGCTTTG CGGTAAGCTG ATCCCAAAGG AAGTAATTCG CATTCGAATC 169401 CAAAGCCCAA CGGTTGTTAT TGCCTTCCGC GGTATGGGCA TACTCATCTC 169451 CTGATTGAAT CATCGGAATG CCTTGCGAGA CCATCAAAGT AAGGAAAAAA 169501 TTTCGTAACT GTCTTTCACG AACTTCAAGA ATGCCAGGGT CTTCTGTTTT 169551 CCCTTCCGTT CCGAAATTGT AGCTGTAGTT CGCATCTGTG CCGTCACGAT 169601 TATCCTCTCC GTTAGCCTCA TTATGTTTGT GGTTATAAGT CACAGTGTCA 169651 CATAACGTAA AACCATCATG GCAACTGACA TAGTTAATCG AATTTGTAGG 169701 CGAGCCGTGA GGATAGATGT CTTGAGATCC TGAAATTCTA GAAGCAAAGG 169751 TTCCTATGAG ATTTTGATCC CCATTAAGAA ATGCTTTCAC GTTATCACGA 169801 TACGGGCCGT TCCATTCACT CCATCTTGGA GACAGTGTGG GGAAATAGCC 169851 CACCTGATAC AAACCGCCAG CATCCCAAGG CTCAGCTATA ATCTTTGTGC 169901 TCGCAAGTAA AGGATCAAAA GAAATCGCCT CTAAAACAGG AGCGAATTGT 169951 AGGGGAGATC CCGAAGGACC ACGAGAAAAG ACAGAAGCAA GATCAAATCG 170001 GAACCCATCG ACATGCATTT CTTCTACCCA ATAACGTAAG ATGTCGAGAA

170051 TCCATTGGGT CGTGGGGGCG CGGTTTGTAT TGAGAGTGTT TCCACAGCCT 170101 GAATAATTTG TAAAGTGACC TTGTGCATCT AAAATATAAT AGCTCGGAGT 170151 GTCTATCCAA GGCAAAGAGC AGGTCGTCCC TTGCAAGCCC GTATGATTAA 170201 AAACAACATC AAGAATGACC TCAATACCTT CTTGATGCAA GGTCTTTACT 170251 AAAGTTTTAA ACTCTCTACT TGGAGCGCAA GGATCAGAGG CATAAGCATA 170301 ACGTCGGCAA GGAGAAAAGA AATTTAGGGG AGCATAACCC CAATAATTGC 170351 ACAGATAAGG GAATTTCGAA TTTCTAAAAG GATGCGCAGT CTCATCGAAC 170401 TCAAAGATAG GTAAGAGTTC AACAGCGTTG ATTCCCAGCT TATGCAGATG 170451 GTCGATCTTT TCAATGATTC CTAGGAAGGT TCCCGGAGCA TGAACCCTAG 170501 ATGAAGAAGA TTGCGTGAAG GAACGTACAT GCATCTCATA GATGATCATC 170551 TCTTCTTTCG GCAAATGCAG AGGCTGATCA CCATCCCAAG GAAATGGTTC 170601 TTCCTTTAAA TAACAAAATG CATAATCCCC CTGTTTCTTT CGCGAACCAA 170651 AACTCTGTGG GGAATGAATA TTCTTCGCAT AGGGATCTGC AAGATATTCT 170701 TTAAAAGAGT ATTGCATTCC ATGCTTTTTA GGCCCATGAA CACGAAATGC 170751 ATAAGACGAT TGATCAGAAA TACCCTCGAT CTCTATATGC CAAATCGCAC 170801 CCGTGCGGTG TGTATCGGGG TAAAGAGGGA CTTCTATGAC TTCTGAATTT 170851 TCGTCTGTTA AAGCAAGGAT GACTTCGGTA GCTTGTGAAG CATATAAAGC 170901 AAATCGATAG CGGTTTGGGG AAATTTTAGA AGCCCCAAGA GGTAAAGGAA 170951 CTGAGGGATA AGAAGAAACT TTTTCCATCG TCGATCAGAT ATCATAACGG 171001 CGCCAGAAAT ACAAGGGAGT TTTTGATTTT AAGCCTCTGA AATCGTTTAC 171051 TAACAAATCA CTCTTATGGT AACTTGAAAC AACAACAATC TATTACAAGG 171101 AGATTCCCTC ATGTCCAGGC AAAATGCTGA GGAAAATCTA AAAAATTTTG 171151 CTAAAGAATT AAAACTGCCG GATGTAGCTT TTGATCAGAA TAACACGTGC 171201 ATTTTATTTG TTGATGGTGA GTTCTCTCTT CACCTTACTT ATGAAGAACA 171251 TTCTGATCGT CTTTATGTCT ACGCTCCTCT GTTAGATGGA CTCCCTGATA 171301 ATACTCAGAG GAAATTGGCC TTATATGAGA AATTATTAGA AGGATCTATG 171351 CTTGGCGGTC AGATGGCTGG AGGTGGAGTA GGTGTTGCTA CTAAGGAACA

171401 GCTCATTCTC ATGCACTGCG TTTTGGATAT GAAATATGCT GAAACGAATT 171451 TACTAAAAGC ATTTGCTCAA TTGTTTATTG AAACTGTTGT TAAGTGGCGT 171501 ACCGTGTGTG CTGATATTTG TGCTGGTAGA GAACCTTCTG TAGATACAAT 171551 GCCACAAATG CCTCAAGGCG GTGGTGGAAT GCAACCTCCT CCTACAGGTA 171601 TTCGTGCTTA ATTAAGCATA TAAGTACTAC ATCTTTTTTC TAAATTTAAA 171651 AGATTGATAA AGCTCTTCTT AGAGAAGAAA TGACTTTATT GACTTCTCGT 171701 TTTTGTCCTG AATCTGTTTT GAGAAGTTCT TAAGATGTAT TAATTAAAGA 171751 AGTATTAAAA ATAGAAATCT AAAGGCTATC TTATGATGTT TGGGCATTTT 171801 GCTGGTTACC TTGGAGCAGA TCCTGAAGAG CGAATGACTT CCAAAGGAAA 171851 ACGTGTGATC ACTCTGAGAC TGGGAGTGAA GACTCGAGTT GGAATGAAAG 171901 ATGAAACTGT TTGGTGCAAA TGCAATATTT GGCACAATCG CTATGATAAG 171951 ATGCTTCCTT ACTTGAAGAA AGGCTCAGGA GTCATTGTTG CTGGCGATAT 172001 CTCTGTAGAG AGTTACATGA GCAAAGATGG TTCACCGCAA TCTTCTTTAG 172051 TGATTAGTGT AGATTCTTTG AAATTCAGTC CTTTCGGTCG CAATGAAGGC 172101 AGCCGTTCTC CATCTTTAGA AGACAATCAT CAGCAAGTGG GATATGAATC 172151 TGTATCCGTA GGGTTTGAAG GTGAAGCACT GGACGCAGAA GCTATTAAAG 172201 ATAAAGATAT GTATGCTGGT TATGGTCAAG AACAGCAGTA TGTCTGTGAA 172251 GATGTTCCTT TTTAATTCCT AGTCATTAAA GGAGAGTTTG TGGTTTTATT 172301 TCATGCTCAA GCCTCTGGGC GTAATCGTGT TAAGGCAGAT GCTATAGTCC 172351 TGCCCTTTTG GCATTTTAAG GATGCAAAAA ATGCAGCTTC TTTTGAAGCC 172401 GAGTTTGAAC CCTCGTATCT CCCCGCTTTA GAAAACTTTC AAGGAAAAAC 172451 CGGGGAGATT GAACTCCTTT ATAGTAGTCC TAAAGCTAAG GAAAAACGCA 172501 TTGTCCTCTT AGGCTTAGGG AAAAATGAAG AGCTCACCTC TGATGTTGTT 172551 TTCCAAACCT ATGCGACACT AACTCGTGTC TTACGTAAAG CAAAGTGTTC 172601 CACAGTCAAT ATCATCTTAC CTACAATTTC TGAATTGCGG CTTTCTGCCG 172651 AAGAATTCTT AGTGGGGTTG TCCTCAGGAA TTTTGTCATT AAACTATGAC 172701 TACCCACGTT ATAATAAGGT AGATCGTAAT CTTGAAACTC CTCTTTCTAA

172751 AGTCACGGTT ATCGGTATCG TTCCCAAAAT GGCGGATGCT ATCTTTAGGA 172801 AAGAAGCAGC CATTTTCGAA GGCGTATATC TCACTCGAGA TCTTGTGAAC 172851 AGGAATGCTG ATGAAATTAC CCCTAAGAAA TTGGCAGAGG TTGCTCTGAA 172901 TCTGGGAAAA GAGTTCCCTA GTATTGATAC TAAGGTCTTG GGAAAAGATG 172951 CCATCGCCAA AGAGAAAATG GGACTCCTAT TGGCTGTTTC CAAGGGTTCT 173001 TGTGTGGATC CACACTTTAT CGTTGTCCGT TATCAAGGAC GTCCTAAGTC 173051 TAAAGATCAC ACCGTCTTGA TAGGGAAAGG GGTCACTTTT GACTCTGGAG 173101 GTTTAGACCT CAAGCCTGGA AAATCCATGC TTACTATGAA AGAAGACATG 173151 GCAGGTGGGG CTACAGTCCT CGGGATTCTC TCGGCGTTAG CAGTTTTAGA 173201 GCTTCCTATA AATGTCACGG GGATCATTCC TGCTACAGAG AATGCTATCG 173251 ATGGCGCCTC CTATAAAATG GGAGATGTCT ATGTAGGAAT GTCGGGGGCTT 173301 TCTGTTGAGA TTTGTAGTAC CGATGCTGAG GGACGTCTTA TCCTCGCTGA 173351 TGCGATTACA TATGCTTTAA AATATTGTAA ACCGACACGT ATTATAGATT 173401 TTGCAACTCT AACAGGAGCT ATGGTAGTCT CTCTAGGAGA AGAGGTTGCA 173451 GGTTTCTTTT CCAATAACGA TGTTTTAGCT GAAGATCTTT TAGAGGCGTC 173501 AGCCGAAACC TCCGAGCCGT TATGGAGACT TCCTCTAGTT AAGAAGTATG 173551 ATAAAACATT GCATTCTGAT ATTGCTGATA TGAAAAATCT AGGCAGTAAC 173601 CGTGCAGGGG CTATTACAGC AGCATTATTC TTGCAGAGAT TTTTGGAAGA 173651 ATCTTCGGTA GCTTGGGCAC ATCTTGATAT TGCAGGTACT GCATATCATG 173701 AAAAAGAAGA AGACCGTTAT CCAAAATATG CTTCAGGTTT TGGTGTTCGT 173751 TCTATTCTTT ATTACTTAGA AAATAGTCTT TCTAAGTAGT TGCTTTCTAT 173801 TTATTTATGT TTTAGTAATG ACTTTTATTT TAGTTTTTTT AAAATAAAAG 173851 TCATTTTTT TATTAAAGTT TTCAATCGTC CCTGCCGATA GATCAGGTAA 173901 GTAATTACCT GTCTAATTAG GGGAATAAAG ATGATTGGAG CGCAAAAAAA 173951 GCAAAGCGGT AAAAAGACAG CTTCAAGAGC TGTACGGAAG CCTGCTAAAA 174001 AAGTTGCGGC TAAACGTACG GTTAAAAAAG CTACTGTTCG CAAAACCGCT 174051 GTAAAAAAC CTGCAGTTCG TAAGACGGCT GCTAAAAAGA CAGTAGCAAA

174101 GAAGACTACA GCTAAGAGAA CAGTTCGTAA GACTGTTGCT AAGAAGCCTG 174151 CAGTTAAGAA AGTTGCTGCT AAACGTGTAG TAAAAAAGAC AGTAGCAAAG 174201 AAGACTACAG CTAAGAGAGC GGTTCGCAAG ACTGTTGCTA AGAAGCCTGT 174251 AGCTAGAAAA ACTACAGTGG CTAAAGGTTC TCCTAAGAAA GCTGCAGCCT 174301 GTGCTTTAGC ATGCCACAAA AACCATAAGC ATACATCTAG TTGTAAACGT 174351 GTCTGTTCTT CAACAGCTAC GAGAAAGCAT GGCTCTAAAA GCCGTGTTCG 174401 TACAGCTCAT GGCTGGCGTC ACCAACTGAT CAAAATGATG TCTCGATAAT 174451 TTGTGATTTT CGCATTATTG CTCATGTTAA CGGGAAAGGG AAACATTGGG 174501 TTTCCTTCCC GTTTTTCTTT TTAAGGTTAA AAAGCTTTAT AGAGCGAGAT 174551 CTTCAGGCTT CATGCTGTAC AGTTGGTAGG AAAATACGTA TAGTAGGTTC 174601 AGGATACTAC TTTTTTGACT CTACCTATGC AAAAATCCTT AACGAGTTTT 174651 GATGACTTTT CCCAGGCGTA TGCAGAGAAA GTGCCCGCTA TAGCTCTTAT 174701 AGGGAGTGCT TTGGAAGACG ATAAAGATGC GCTGATTGAA TTATTAGTCT 174751 CTGAGAGCTT CAAAGAGCTC GGTGGTCAGG GACTCATGCC AGCAACCCTC 174801 ATGTCTTGGA CCGAGACGTT TGCACTCTTT CAAGAGCATG AAACTTTGGG 174851 GATTATTCAT GCAGAGAAAT TCCCTCTAGC AACTAAGGAA TTTCTAAGCC 174901 GCTATGCTCG GAATCCTCAA CCTCACCTTA CGATTTTGAT CTTCACCACA 174951 AAACAAGAAT GCTTTCGAGA ACTGTCAAAA GCCTTGCCAT CGGCTCTTTC 175001 TTTGAGTTTA TTTGGTGAGT GGCCCGCAGA TCGTCAGAAA AGGATCATAC 175051 GCCTCCTGTT GCAAAGAGCT GAGCGTGTGG GGATTTCTTG CTCTCAATCA 175101 TTGGCATCTT TGTTTTTGCG TGCACTTGCT TCAACCTCTC TTCCTGATAT 175151 TCTCAGTGAA TTCGATAAGC TACTGTGCTC TGTTGGCAAG AAAACGTCCT 175201 TGGATCACTC TGATATTAAA GAGCTCGTTG TCAAAAAAGA AAAGGCTTCC 175251 CTATGGAAAT TTCGAGACTC TCTATTGAAG AGGGATCCGG TAGAAGGTCA 175301 CCAGCAGTTG CATTTTCTAC TCGAGGATGG TGAAGATCCC TTGGGGATTA 175351 TTACTTTCCT TCGTACCCAA TGTCTCTATG GTTTACGTAG TATTGAAGAG 175401 GGATCGAAAG AAAATAAACA CCGAATGTTC GTCCTTTATG GAAAGGAGAG

175451 ACTACACCAA GCCTTAAATT CTTTATTTTA TGCAGAAACC TTAATTAAAA 175501 ATAATGTCCA AGATCCTATA GTAGCTGTGG AAACTTTAGT GATTAGAATG 175551 GTTAACCTGT GACTTTATAT CTTCTTCCCA ATACTCTCGG TACCCGTGCT 175601 GTAGAGACTC TCCCCTCCGT TATAGGAGAA TTAGTTCATA GACTAGATGG 175651 GCTGATTGTA GAAAGTGATC GTGGGGGTAG GGCATTTCTA AGTTTATGGA 175701 AAATTCCCGA AGTTCATAAA TTTCCTCTTG CTATTCTTAG TAAACATGCG 175751 CGCCTCCCTA AGGCTTGGGA TTTTTATCTA GAGCCTATCG TAAAACACGG 175801 GGAGAATTGG GGACTGATCT CTGATGCGGG TCTTCCCTGT ATTGCAGATC 175851 CTGGAGCGAG TTTAGTGCGT CGTGCACGTG CTTTGGGGAT TCCTGTGCAG 175901 GCTTTTTCAG GTCCCTGTTC GATAACGTTA GCGCTCATGC TTTCAGGCTT 175951 GCCTTCCCAG AGCTTTACGT TTTTGGGATA CCTCCCGCAA AGTCCTAAGG 176001 AACGTGTAAA GTCGATAAAA AAGGCAGCGA CCTCCAAAGA GGTATCTACT 176051 TCAGTATGTA TAGAAACTTC TTATCGTAAC GTCTATACTT TTGAGTCTCT 176101 TCTAGATACT TTACCTTCCT ATGCGGAGCT TTGTGTTGCC TCTGACCTTT 176151 CGGGCCCGAG TGAACTTGTT CTCACACGCC AAGTGCAATC ATGGAGAACT 176201 ACTGAGGACT TAGGTTCTGT GAAGCAAAGT ATAACCAAAG TCCCAACAAT 176251 ATTTCTTTC CACATCCCGA ATTAATTTTT TCTTAGCTTG TTTTATCCCC 176301 GGTATCATTG TCAGGAGTTT CTTAGCATGT AAGCTCTTCG GGGATATTTT 176351 TCTTGTGATC ATAAGCAAAA ATTAAAGGTT GCATGATTCT AAATAGCGCA 176401 TAATCATAGG TTTGCGTCTT TTTAAAATAA TCGTAATCCT CTATTGCGTG 176451 TTATCAATAC TTATTTTATA ATCTTTGCCA AATAAGTTTA GAGCGTTGAT 176501 GACAGCACCC ACAGAATCTC GATCTTCACC ACCAACACTA CTCGAAGAAA 176551 CGGAACCTCT ATCCCCAAAT CCTATTCCCG CCGATATCCA AATCCCGAGA 176601 ATTACAATAT CTCCTCCTTC TCTAGACGTA TCAACAGTTG CATCTTCGGC 176651 TGAAGATATC TCGGTCTTCA TTGCAGGAGG CCCTAGAAGT TCTTCATCTG 176701 CTTCTGTGGC TTCGGATGTG TATGAGTTGG TTTGCCTTTG TGGCGGTGAT 176751 GAGGATCCTG AGCCCCCAGA TTCTGAGGTA CGTACTCTCT ATGTAAATGG

176801 AAGTTGGCAA ACGCATCAAG AAGCTGTACA GGAATTGCTG TATATTAGTG 176851 AAGTACGGGG AGAGGCCGTT CGTCTTCTCT ATAACGATGG AAGCGGCATG 176901 TCTCCTTGGC CCATCAGTCC TTGTCGTACT CTTCCTACTC TCGATCATCC 176951 TTTATGTCAG GCCCTTTTGA CGGTTTGGGA ACAGTTTTTC TCTGCTCCTG 177001 AAAATCAGAA TCGTGAGTTT CTAGTGATTT TCTATGGGGA TGCATCGCCT 177051 TATATACAAC AGGCGTTAAC GCAATCTAGG CATAGTCCAC GTATTGTTGT 177101 TGTAGGGATT TCCCCGACGG TCTTTATTCA AGGAGACTTT AGGGTCCATA 177151 ATTACCGTGT TTCTGGAGAC TTCTTTAGTT CTCTGGATTG TCGGGGAACT 177201 AGGGCGGAGA ACACCACGAT ACTGCCGTAT TCTTCGGGTC TTGAGGGTGT 177251 TTTTCTGCCT TCTATCCGTT GTCCTTCTTT TACTTGGGCG GTGCGTTTTG 177301 GAGAGCAATG CTTGGTTGCG AATAGGGGTG AGGATGTAGA AGATAGGGGA 177351 GGTCTTTCTC AAGATGCCGA AAGATCACAG TTACCACACA GTGAAAGAGA 177401 TCTAGCTGTT GTCATTGATT CTACGGATCC TAGTTCTATG AGTAGGCTTG 177451 TAGAATGGTT GAATCAAGGA TCGCCTTCAT CAGATATGGA AATCAATCCC 177501 TATCCCCAAC GGTGTCCTGA TGTAGCTCTT TCTGCGCTTT ATGCAATTTC 177551 TAGAGTTTCA GGACTTGCGC AGGAATGGAT CCTAGCCTCT GTTCATGAGG 177601 GCTTAGACTT GCAGATCTGT TACTCTTTAA TTTTGATGCA CACGACGTTT 177651 GCGGTCCGGT ATTTTTCTT ACTCTTTACA AATTATCCTC AGTCTAGAGA 177701 GAGATTCCGT ACTGCACGAA TCGTAGCACA ATCTCTATAT TTACCAAGCA 177751 TCCTTGTTCT TGTTTTTGAT TGTGGCAACG TCCTGCGTAA ACTATGGATG 177801 CCTCAGGAAA TCTTACGAGC AATTTTATT TCTGCGTCTA CAATTTCAGG 177851 GAGTATTGTC TTTGTAGAGT GCACTCGCTG GATGGGGCGA GGTCTTAGAC 177901 ATCGTGTACA ACAATTTGTG CAGCAACGAG TTATAGGAAG TGGCCTGCCT 177951 GTAGGAACAG TACGAGCTTC TTATCGCGAT CGTGCAGGCT TTATCATAGG 178001 CTTTTTACAA ACTGTACATG GAGGACTTTA TTTGCCGGŢA TCCATTATGG 178051 TGCTTAACCA GATTGCAATA CAAGTTCCAC GTATCTTAGT ACGTCCAAAT 178101 AACACTGCTG TTTATGATCT ACATAATAAA AGTGCTGAAG AAAATTGGAG

178151 CAGTGGTGAT GTATTAGCTG TTGGCCAAAC ATTAAACTTC ATCTTATGTG 178201 CTTTCGTCTT GTTCGTAAAT CTATGGTTCT TTGTGAAGTC CGTATTACGC 178251 CATTCTAGGC GTCGTCGTCG CTAATGTATC TTTGGAGACA TCTTGGTATT 178301 TTTAACATAA GAGCACGGCA AAAAAAAGAA AGACCGAGAA AGAAGCTTTA 178351 CCTTTAAGAG TTCTCTGATG TTTCCCGGAG CATCAGGAGT CCGAAGGGCA 178401 GAGTCTAGGT TTTAGCCTTC CGCAGGGATT AGAAGGAATA TCCCATAATC 178451 TCTTCCGCTA TTGTATCGTG GAAGAGACGG CCCTGTCTAT TTAGGGCAAG 178501 ACATTGTCCA TGCACACTGA ATAGGTTTTG TAATTTTACA TCTTGCGTAA 178551 GCATGGAGAT AAGTGTGGAG GGGAACTCCG CGAGGTCTGC TCCTTCAAGG 178601 AGTCGGAGTC GCAGGGCTAA GGCTTCTTTG ATTCGTTCTT TTTTTGGGAG 178651 AATTTCTGAG GTCTCTTGGG TAGGGAGATT CTTACGTACA GCACGTAGAT 178701 AGTGAGAAAT ATGACTATAA TTTTTTGACC GCTCTCCGTG AAGGTATTGC 178751 GAAGCTGAAA CTCCTAAGCC TAAGAAAGGG ĆGATCTGTCC AGTAATAGAG 178801 GTTGTGCTTT GCGGGGTAAT CTGGCTTGGC ATATGAAGCA AGTTCATAGC 178851 GTTGGAACCC TTGGGAGAGT AGGAGATTTT CAGCAAGGAG GCTCATCTCA 178901 GCTAGAATTT CTTCTTGGGC AATTGTGGGG ACTAGAATTT TGCGGTGTTT 178951 ATAGAAGGAG GTGTGGGGAT CTATAGTGAG GTTGTATAGA GAAATGTGAG 179001 TGATAGGGAG AGTCAGAGCT TGATGTAGGT CGCTTAGGAA TATCTCCAAA 179051 GACTGTGTGG GCAGTCCGTA GATTAGGTCT ATAGAAAGAT TAGAGAATCC 179101 GTGATTCTGG CATTCTTGCA GTGCTGTGAT TGCCGCAGAT GAAGAATGCG 179151 TTCTTCCGAG GAGCTGTAGG ATAGAGTCGT CGAAGGTTTG TACGCCAACG 179201 CTAATTCTAT TTATTGGAGT CTCTTGTAGT TGACGTAGAT AGCTTACGGT 179251 GAGATTTTCG GGGTTGGCCT CTAAAGTAAT TTCCCGGGCA TGGGGGGCTA 179301 GCTCTTTGAG GATGCGCTTA AGATCAAGAG GAGAAACTAA TGAAGGTGTT 179351 CCCCCTCCAA AAAACACAGT CTCTATGAAA TGCGTCTCTT GGATGGGGGC 179401 TAGCTTTCTT AGCCCCTCTT GAATTACAGC ATTACAATAG AGCGATACAG 179451 ATTCACTTTT GTAGGGGATT GTATAAAAAC TGCAATAGCG ACATTTTTTT

179501 GTGCAGAAGG GAATATGAAT ATAAAGAGCT AGGGGAGCCT TACCATTCAT 179551 TAGCATCGGG ATCAATGATA CCACCGCGTC TCCAGCGATT GCGATCGCTG 179601 AAAGGGTCTT CATCGTCTTC GTGGGAATAG TCGACTTCTA CTGCCCCATC 179651 TTCACGTCCT TCATCGGTGA GTTCCAGTTC TACGGTTTCG CCTGGGTCGA 179701 TATCAATTTC TGTTTCTGTG GGTTCCACAA ATTCAATATC AGACATACTA 179751 GGACCATCAG GGTAGACAGT CTTTGCGGGA CTGCGTCGTG GTGTGACGTA 179801 CTCTTCAATC TCTCCGTTTT CTTTGAGGAG TTGTAGACGT TCTTTTTCTT 179851 CGTAAATTTT ATGTTCTAAG ATTCGGATTT CTTCTTGGTG CCTGCTAATT 179901 TCTTTTTTG GCACTAAACC CAACTGCATC CATTGCGTAA GGTCATGAAG 179951 TTCTGATTCT AATTTTTTAA GACGTTCGCT TTTCATCCAA GGGTTCCCAT 180001 CCTGTATTCG AGAGCAGGTG CATAGCATTT TTTTGTTTTT TCAACTGCAG 180051 AAAAAAACA AAGATTTTTT TTTTTTTGAA AAATAAAAAT TTGCTATGAA 180101 GCAGACATTT AGATCGTATT TATTGAGTTT AATTATTTTA TGGATTCCGA 180151 GTTTGTGGGG CAAGTATATT CTTCGGATAT GGATTGGATC GAGTCTATGT 180201 ATCAGAGATT TATGAATCAC GAGACTTTGG ATCCTTCTTG GAAGTATTTT 180251 TTTGAAGGGT ATCAGCTCGG TCAAGCAGCA TCTCCATCAG AAGCTAGTAC 180301 TAAGATTTCT GGGAATGAAA CTATTGCTAT GCTTCAAGAA CAAAAATCTC 180351 AGTTTCTATG TACGATTTAT CGTTATTATG GATATTTGCA AAGTCAAATT 180401 TCAACGCTTG CCCCAACTAC AGATTCTCGA TTCATTCAGG AAAAGATCGC 180451 TAAGATTGAT CTGGATGAGC AGGTGCCTTC TGCGGGTCTA CTTCCTAAAG 180501 CTCAGGTTTC GGTACGAGG CTGATCGAAG CTTTAAAAAA ATGCTATTGC 180551 GGAAGTCTTA CTTTAGAAAC CCTAACATGT ACTCCTGAGT TGCAGGAGTT 180601 TGTTTGGAAT CTTATGGAGA AGCGACAAGT GGAGCGCTTT GCAGAGCAGC 180651 TCCTTCGCTC CTATAAAGAC TTATGTAAAG CAACGTTTTT TGAAGAGTTC 180701 TTACAGATAA AATTTACAGG TCAGAAACGT TTTTCTTTAG AGGGCGGAGA 180751 GACCTTGGTC CCCATGTTGG AGCATCTTGT TCATTATGGA TCGGCATTAG 180801 GAATTTCTAA CTACGTTTTA GGAATGGCCC ATCGAGGTCG TTTGAATGTA

180851 TTAACGAATG TTTTGGGAAA GCCTTACCGT TATGTCTTTA TGGAGTTTGA 180901 AGACGATCCT GCAGCACGTG GTTTAGAGAG TGTTGGGGAT GTAAAGTACC 180951 ATAAAGGGTA TGTGCTAAAG TCCCATCAGA AAGATAGGGA AACTACCTTT 181001 GTGATGTTGC CAAACGCTAG TCATCTCGAA TCTGTAGATC CTATTGTCGA 181051 GGGGGTCGTG GCTGCCTTGC AACACCAAGG TCACGCAGGT AAAGAGCAAA 181101 GCAGCTTAGC AATTTTAGTT CATGGAGATG CAGCATTTTC TGGTCAGGGA 181151 GTGGTTTATG AAACTCTCCA GCTGAGTCGT GTTCCAGGGT ATTCTACTGA 181201 GGGTACGCTT CACATTGTTG TGAATAATTA CATAGGGTTT ACCGCAGTGC 181251 CACGGGAGTC AAGGTCCACC CCTTATTGTA CGGATATTGC TAAAATGCTA 181301 GGGATTCCTG TATTTCGAGT GAATAGCGAG GACGTCGTTG CCTGTATAGA 181401 TCATAGATCT CTGCTGTTAT CGCAAGTATG GACATAATGA AAGTGACGAT 181451 CCCTCAGTAA CAGCTCCCTT ACTCTATGAT CAGATTAAGA GAAAGAAGAG 181501 TATTCGCGAG CTGTTTAGGC AATATCTGTT GGAAGGGCAG TTTGCAGATA 181551 TTTCTGAAGA AACTTTGGCA TCTATTGAAA AAGAGATTCA AGAGAGTCTG 181601 AATCGTGAGT TTCAAGTATT GAAAGGGACG GATCCAGAAC CCTTTCCTAA 181651 AAAAGAATGT CATCACTGCG ATCGCTTAAA TAACGGCGAG CTTATTTTGC 181701 ATGATTGTGA TGTTTCTTTG GATCGCGAGA CTCTTTTTCA TATGAGCTCG 181751 CGTCTTTGTG GTTTCCCTGA CAATTTTCAT CCCCATCCTA AAATTAAGAC 181801 TCTTTTAGAA AAAAGAATGA AAATGGCAGA AGGTGGGGTT GGTTATGATT 181851 GGGCGATGGC CGAAGAATTA GCCTTTGCTT CGCTATTAAT CGAAGGGTAC 181901 AACCTGAGAC TCTCAGGTCA AGATTCTATT CGCGGGACAT TCAGCCAACG 181951 ACATTTGGTA TGGAGTGATA CTGTGACTGG AGATACCTAC TCTCCATTGT 182001 ACCATCTTTC TGCAGAGCAG GGCTCTGTAG AAATGTATAA TTCTCCTCTT 182051 TCCGAATATG CAATTTTAGG GTTTGAGTAT GGCTATGCTC AACAGGCATT 182101 AAAGACTTTA GTGTTATGGG AAGCGCAGTT TGGGGATTTT GCTAATGGTG 182151 CACAAATCAT TTTCGATCAG TATATCTCTT CGGGAATTCA GAAGTGGGAT

182201 TTACACTCTG ACATTGTTCT GCTTCTTCCC CATGGGTATG AGGGCCAAGG 182251 ACCCGAGCAT TCTTCATCTC GTATAGAACG TTATTTGCAA TTAGCCGCGA 182301 ACTGGAATTT TCAAGTGGTC TTGCCTTCCA CTCCTGTGCA ATATTTTCGG 182351 ATTCTCAGAG AGCATGCTAA GAGAGATCTT TCTTTGCCTT TGGTGATCTT 182401 TACTCCTAAG TTGCTGCTGA GATATCCACA ATGTGTAAGT AGTATCGAGG 182451 AGTTCACAGA ACCTGGGGGA TTCCGTGCTA TTCTCGAAGA TGCCGATCCT 182501 AATTATGATG CTTCTATTTT GGTATTGTGT TCGGGAAAGA TCTATTATGA 182551 TTATGCAGAA ATGCTTCCTC AAGATCGGCG TAAGGACTTT TCTTGCTTGC 182601 GTATAGAGAG CTTGTATCCT TTAGCTCTTG AGGATTTAGT GAGCCTTATC 182651 GATAAGTATT CTCATTTGAA ACATTTTGTT TGGCTACAAG AAGAATCCAA 182701 GAATATGGGG GCCTATGACT ATATGTTTAT GGCGTTGCAA GACATTCTTC 182751 CTGAGAAACT GCTATATATA GGACGTCCTC GGAGTAGTTC CACAGCTTCT 182801 GGATCAGCGA AGCTCAGTCG TCAAGAGCTG GTCACGTGTA TGGAAACCCT 182851 CTTTTCTTTA AGGTAAATTA TGACTACAGA AGTACGCATT CCTAATATTG 182901 CAGAGTCGAT TAGCGAGGTG ACCGTAGCTT CCTTGTTAGT TACAGAGGGT 182951 GCTCTGATTC AAGAAAACCA GGGCTTACTA GAAATTGAAA GTGATAAGGT 183001 AAATCAGCTC ATTTATGCCC CAGTATCGGG AAGAATTTTC TGGGAGGTTT 183051 CAGAAGGCGA TGTTGTTCCT GTAGGGGGGG TAGTGGGAAA AATAGAGCCC 183101 GCAGGTGAAG GGGAAGAGCT TGGAGATTCT CAGTCTAAAG AGACTATAGA 183151 AGCTGAGATC ATTTGCTTTC CTCAGTCTGG GGTGCGTCAG TCTCCTCCAG 183201 AGAATAAAAC GTTTATTCCT CTTCGTGATC AGATGGACCA AGGATCCCAA 183251 GGTCTTTCTG CAGGAGATCG AGGAGAACT CGAGAACGCA TGACCTCGAT 183301 TCGTAAGACA ATTTCGCGGC GTCTTTTGTC TGCTTTACAT GAGTCTGCGA 183351 TGCTCACGAC ATTCAATGAG GTCTATATGA CACCTCTTTT TCATTTGCGA 183401 AAGGAAAAC AAGAAGAGTT TCTATCTCGA TATGGGGTGA AGTTAGGATT 183451 TATGTCTTC TTTGTGAAAG CTGTCTTAGA GGCTTTGAAG GCATATCCAC 183501 GAGTGAACGC CTATATTGAT GGCGAGGAGA TTGTTTACCG TCACTATTAT

183551 GACATTTCTA TTGCTGTAGG TATCGATCGA GGACTTGTGG TTCCTGTGAT 183601 ACGCGATTGC GATAAACTTT CTAACGGGGA GATTGAGCAG AAACTCGCAG 183651 ATCTTGCCCT TCGGGCTCGT GAAGGCCTAC TTGCAATAGC GGAGCTTGAG 183701 GGAGGAGGTT TCACAATTAC CAATGGAGGC GTATATGGAT CGCTACTTTC 183751 GACTCCCATT ATCAATCCCC CGCAAGTGGG GATTTTGGGG ATGCATAAGA 183801 TAGAAAAGCG CCCCGTTGTT CTTGATAATG AAATTGTAAT TGCAGATATG 183851 ATGTATGTCG CTTTAAGCTA TGATCATCGT CTTATTGATG GGAAAGAGGC 183901 TGTTGGGTTT TTAGTCAAAG TGAAAGAAGG CCTAGAGAAT CCTGCCTCAT 183951 TACTCGACTT GTAATTTCTC TGATTCTCAT AAAGGCTCTT TTAGAGCCTT 184001 TTCAGATTTT TTAACCTCTT TTCTTATCAT GAAAAGGATT GCACTCATCT 184051 TGAGTAAGGA ACAATGTAAA TTGAGTATCC GCACTTACAT TGGCTTGTAC 184101 CGCAGCCTCA TTCAAGGCGC CTAACAACTT TCCGGGTAAG TACACATTGT 184151 CTCTTTTAAA GTGAGGTTTC CCGGTATGAT CAAGAGGAGT GAAAGCTTCT 184201 TCTGCGTGAA TATGTACTGA TGTGAAGCCA TTTTCTAAGA GATCTACAAA 184251 CAGATTTTTC AAGAACTCCT TTGAGTCTTT TTCAGAAAGC GAAGAGCTCA 184301 TCGGATCTTT TCTTTCTAAG ATATCTTGTA CATCCTTGTC GTAATAACCC 184351 CAAGGAGTCC ATGCTAAAAC TAAAGTTGAA CTGACACCAG GAGTTGGAGA 184401 ATTGAAAATT GCCCATTGAC CTGCTTTTTT TCTAGAGTTC ATTTTAGTTT 184451 TCATAGCTTC CCACTGTTCC TCTCCTAAGA GATCTTGCAG TTGCTTTAGG 184501 GGAGCCGAGA CTTCCTTATC CCCTTTCCAA TGCGAAGTTT CTGGTGAAAG 184551 GAGGAATAGC GAAGGTACGC GAGCCCCGTG ATCTCTAATT TCTGTGAGAT 184601 TAGGAGCTTT AGCTGTGACA ACCTTAAATT AAGGACCCTT CTTTAGATGG 184651 TGAATTTCAC AATTCGTAGC TTCTGTAATT TTGATTTCTT TAGTGAGTTT 184701 TTCTTTGGGA GGAAGAGCGT CGTAAGGTAA GGAACAATGA TGTTAATAGG 184751 GCTAGTTTAG TTAAATAAGA ATAGGGGATC TCAGGTTTAG GAGTGATGGG 184801 TTGTTGCTCG ATTACAATGA TGGTTAATGG AAGAAGAGCG TTTCTAGAAA 184851 GTGCATCTGC TTCCTCGTCT TCTTTAGGCA ACAAGCTCAA GATGAGAAGC

184901 ACGAGCCCGA TGGCTAATCC CACGAGTGTG ATAATACCAG CAATGCTATA 184951 CGCCTGTATT AGAGGGAAGC CTAAAGCTAA GAAGGTAACG AACGTGATTC 185001 CAGCTGCTAA CATTAGTGCA GCGAAAAGTA TGTAAAGAAC AGCACGCGCA 185051 ATTTGAAGAG CAAGACTTTT ACTTCTGAGT GTGCCTTCTG TAGCTTGGGT 185101 TTTGTTAGCT TCAATCTCAG CGAGTTGCGC TTTTTTTTGGC TTCACTGGTT 185151 CGTGAACTTT ATGATTAGTT GATGCGCCTA ATTTCATAGA AAGTCTTTAA 185201 GCTTAATTTT TTCCTAAAAA TTGTAGTTTA CATGTTGTTT TTCATCAAGA 185251 AGATTACTGA ATTCTGAGTG GGTTGGAACG AAGCATGTTC TTCTAGAAGC 185301 TAAGGTCCTT AGGGGAAAGG AAAAGAAGCT TGGCTGGATC TTTTTAATCT 185351 CTGGGTAGGA GAAGGACGGC GGTTACATTA TTTCTATTTT TTGAATGTTG 185401 GCCTGTGAGA TTGGGATCGG GATGGTGAGC AAGGTACCTT GAAGAAAAGA 185451 AAGCGTCGTG TTCCGTGTAG GTACAGAGGT CTGAGATAAA AATGCGGTCT 185501 TTAGAGATTC CTAAATTCGT AAGTTGCTTG CGAGCAATCG CACGCAGGTC 185551 AAAATGGTTT TTGGGATTCA TAAAGGGAAG AAAGCTACGA GGAAATAACG 185601 TAGCGTAATC GGGATAGATA GCATAATCTG GACCGATGGA AGGGCCGATA 185651 GCTACGAAGA GATCTTGTGG TTTTGTATGA AATAATTTTT TCATAGTACC 185701 TACGGTGACA GCATAGATAT TGCCAAGCAA TCCTCGCCAT CCGCTGTGTA 185751 CATTTGCGAT TGCGTGGTGT TCTCGATCAT AAAAGATAGC TGCTTGGCAA 185801 TCGGAATGGC GGATATGGAG AGAGAGGAGC GGAGACTGCG TGCACAGTCC 185851 GTCTGCAGGT TGGTAGGTGG GGGATGTAGG TGTAACACAA CGTACGGAAG 185901 TGCCGTGGCG TTGATGAAGG TCGCAATACT TCGGAGATTG GAGAGCTGAA 185951 GCAATCTCAG GATTCTTGGC TGCGAAGACC GTGCCCTCGG CATCCTTTTG 186001 TTTTGAAAAG ACTCCGTGGC GTATAGGCAA ACCATCGAAT TCTTCGAAAG 186051 TCCAATAGTT CAGAGGGTGA GTGTGGAAGG ATAGAGTCAT ACTGTCAACT 186101 TAGTTTCTTC AGGGTCTTTC CATACCCCAT GTTCTTGTAA AAGTCGAATT 186151 AATTCTTCTT CAGCATCTTC CATGGGTATG TGAGCTTTTA CACAAGTATG 186201 TTTTACATAA AGATCGATCA TCCCTGTTTT GGAACCTACA AATCCAAAAT

186251 CTGCATCTGC CATTTCTCCA GGGCCATTCA CAATACAACC CATGATAGCG 186301 ATCTTAAGTC CTGGTAGGTG CTGCGTTCTC TTGCGGATAC GTGTGGTGAC 186351 TTCTTCAAGA TCAAAGAGGG TCCGACCACA CATAGGACAG GAGATGTACT 186401 CTGTTTTTAC AAGGCGCACC CCTGCATTTT GTAGAGTGCC AAAGGCAATT 186451 TTTAGCACGT CCTGTAGGGG AAGGTTCGGT AAGTCAAGAA CCACAGCTTC 186501 TCCAAGGCCA TCAAGAAGCA GAGCTCCAAA CTCTGTTGCT ATGGAAATAG 186551 CAGCTTCTTC TTTATTGTCA AAGTCCCTTG AAAATACTAG CTTGGTCGGT 186601 TTTCCTTGGT GTCCTTGTTT TTCAAAGAAA TCTCGGGAGG TATGAATGAA 186651 GGGGTCTGAA GCATGAAAAT GCACAAATGG AGCTTGATGA ACAGCAGGGC 186701 TATCCCAAAT CTCCTCATTG TGTTCATATA GGCAAGGCAC TTGATGGTGG 186751 TGGAAAACTA AAAAGTGTTC TCGAAGTACA TCTGTAATAG GAGCATCTTT 186801 TAACTCAGGG GGAACGACGA CCCCTTCAGG AGTTGTGAAT GCTTTTCTT 186851 TTGTTACGGG ATTTACCCCC AAGTGTTCTA AGAGTTCTTC AGGAGTAAAG 186901 TCGGTAAGAT GGTGAGGATA GAGTTTTAAA AAGACTCCGT AGACGTCTCC 186951 CCAAAGTGTT GTTTTCGCAG GCTTCTCTGC AGCAGAAACA AAGTTTTCGG 187001 AGTGTTGTAG GGAAAAGGGA TTTTTCTTTT CTGGAAGGTC TAAGTAGATT 187051 TTCGTATGGC GTAGCAAGCT ATCACAGACA GGAATTTCTG TAGTGGGACA 187101 CCCTGTGAGA GAGCAGCGTA TGGTATCCCC GAGTCCTTCG GCAAGAAGAG 187151 TTCCGATTCC TACTGCGGAT TTTATGATCC CGTCCACGCC CATTCCAGCT 187201 TCAGTAACTC CAAGGTGAAG GGGATAGAGC CAGCCTCTAG CATCTAAGTC 187251 TTTAGCAAGT TGGCGGTATG CAGTTACCAT GATCTTCGGA TTGCTAGATT 187301 TCATTGAGAA GACAACATCT CTATAATTCA GCTTTTCACA TACAGCGATA 187351 TATTCAATTG CTGAGGCTAC CATTCCTTCG ATAGTGTCGC CATATTTTTG 187401 CATGATTCTT TCGGAAAGTG ACCCGTGGTT CACTCCAATG CGCATAGCCT 187451 TGCCTAGTCG CTTACATTTC TCTACTAAAG GAGCAAACTT TTCTTCAAGA 187501 CGCAGGAGAC TTTGGGCATA GCTTGCCTCT GTATAGATCT TCGTCCCCTT 187551 GAACATGTTC CTCTTATCTA TGTAGTTGCC TGGATTGATG CGAACCTTGT

187601 CAGCAAAATC AGCAACTAAC ATAGCTGCTT GAGGGAAGAA GTGGATATCT 187651 GCAACCAAAG GGATATTTAA CCCTAGAGCA ATCAGACGTT CTTTAATTTT 187701 TTCACAGGCT TGTGCTTCCT TGATTCCCTG TACAGTCACT CTGACAATAT 187751 CACAATTATG TTCCGCTAGA GCGTAGATTT GCTCTACTGT ACTGTCAATG 187801 TCTGTGGTTA ATGTCGTTGT CATTGATTGG GTTTTTATTG AGTGGTCACT 187851 GCCTATGTAT AAGTTGCCTA TTCTTACTGT ATGGGTTTTG CGTCGCGAGG 187901 AATTGATGGC AGGGGTAATG AGTGTCATAA AAATCTCAAA AATTTCAGAG 187951 TTTTATCAAT TATAAAAGCC GAGAGAATTT TGTTGAAGCT GATAAAGCTT 188001 CTCTAAGGAT GCCTACATTT ATCAATTTTA CAGAACTTCA CCTCAATAAG 188051 AGGAATTTTC AGAGAGAAAC GACGACCTTG GATGGCCGAG TGTTTCAGGC 188101 CATCAAGAGA GTATATGGAG AGATTATTGT GAATTAAAAT AATTATTTCA 188151 TACTTATCAA GGGGATAAAG ATTTTGCTTT AGAGGACGAT TCGTAAGGAA 188201 CTCGAATCTG TTTTTGTGTT CGTAGAACGA ATCCCGAAGT TTGACAAAGA 188251 GAAGGAGTTC CCAAGAGGAG TGCTTGGGGC TCGTACTTTG GGAACGCTTT 188301 TTTCTAATTT GAAATCGGTG TATTTTGTGA TGGATGCGTG ATTTTATACA 188351 TCGAGATTAA CATGATAGCC AGAAGGATTA GGGAGAGCAA GAATAAGGGC 188401 CCTATAACTG TAAAGTAGGC TCCGACAGAT GTTGCTACAA AGATCATGGC 188451 TATAGCAACG ATTCCAGAAA CAAAGGTATT GAGGAGGAGT AAAGCAGTTG 188501 AAGCCGCAAA GGCAATTTTT ATGCATCGTC CTGATCCCGA TAACCTTTCT 188551 AGGAATTCTC CAAGCTTTGT TTGTTCAACT GGAGATGCAC TTGATGTTCC 188601 TGTGACTACT GGAGAAGACA TGTGATTTCC TATATTTCTA TGACAAGCCC 188651 TATAATTTTA TAGTTTTCTA TTTTAAATGC AATGTAATTA GTGTTTTACA 188701 AATTAATCTC CTACTAAAAT AGGAGGAGAC ACTCTCATAT TTTCTATGAC 188751 CTAGGGAATA AAATATTTTA TTTATAATGA GTGAAAAACA CCGCTGTTTG 188801 TCTAAGAAT ACTTTTAGAA GAGAGTTTAC GTTTTATCTG CTGTTCAGCA 188851 GTTGTTCCTT CGGTTTTGTT TACAAGGGCA AGCTTTATAA CAAGAAGAAG 188901 CACACAGAGT ACTGTTGTAA GAATAAGGCC TCCTAGGATA ATCAAGCAGA

188951 TGGGAGGAGT CCCAGGAATA AAGAAAGCTG CAATTATTCC TAAAACTTCT 189001 ATTACTAGAA TGGCAACAAG GATTTTAACT AAAAGTTTAA TCGTCTTCGT 189051 CGTTTTATCT TGATGCTTTG CTTTTAGAGC TTTGGAATTT GGTTCCATGA 189101 GGCCTAGATT TCCAGAGGAA GGGCGGTGGC TTGTGGGTAG GGGGGCTGAG 189151 GACACGGGCA TCGTCTTTTA GTTAGAATTT TCTTTATTTC GCAAAAGCTT 189201 TGTCTTATGG AGAGCAGAGG CGTTTCTATT TGCTTGAGAA AGTATAGCGT 189251 GGTTTAGAAA GTTAGAAAAG ATTTTTTAAA AAATCTTCTA TGTTTTGGAT 189301 GTTTTAAAGA GAGAAGTCTT ACCATTTTTT AAGAAAGGAT TTCCTTTTCA 189351 CAAAAAGGCA TCTACTCTCT TTTGCTTAAC CAGGGGGAGA GATTAGGGTG 189401 TGATGAGGGG ATGTCTAGCA AGAGTTGTCA AAATGATAAC CCACGGTTGA 189451 TTTTGATTTT CTGCTTCTGA TCCAAATGTC TGCAGAGCAT CTATCAAGGC 189501 TAACTTCACA CCATCAATCC ATTGCAGACG TAGACTAGTA GTCTCTTCGT 189551 CTTTTGGAGA GCCAGGGAAA TGGGAGGAAA GCAAGGGGAG CTGAACTACG 189601 GTTGTTTTAC GGCGCTTGGC CTCATTCAAA CAGTTAAGGT AGGCTTGCCT 189651 ACAAAAGGTA AACGCATCAT TAGGATTGTA GTTATAGTCA GAAGCTTTCG 189701 GCCCTAGAAG TTGTCCTAAG AATTGCGGTA GTCCTGCTTT ACCTGCATTC 189751 GTGGTTCCGT TTAGATTTTC CCATTTTGCT GAGCGGCATT CTCCTAATTG 189801 TAGGGGCTGT TTTGAACGGA GAGGATCCGG AGATTTTTCC GAATTTTCCC 189851 AACAGCGTAC ACTAGTTGCT TTCGCAAGTG CGAGACTTGT TCCTTTTACG 189901 TTGTTTGCCA TGTTTGGGGT GGCTGCATTC ACAATCATGA CGATGCCTTG 189951 AGTTTTATAG CGAGGCTTTT CAATGGGCCC TAACGTAGAG ATCAAAGTCA 190001 GGTTTGAGTT TTTCAAATGC CAAGCATAAC CGGTTTGATT GTCAGCAGCG 190051 ATGAAAGGCA TATCTGTGTT AGCCTGTAGA TCAGGTATAC GGTCCCAATT 190101 TTCTTGTAAC AGTGTTTGAT GGGTTCTAGG GGAAGGAAGC GGGGGCTCTT 190151 CTAGGACAGC TTCCGTAGGC ACTGGGGTGA GTGCTAGGGG CGGGACAGGA 190201 GGAAGGTCTG TATCTTTTAT GATGGGCGTC GGCTGATGCG TTACGCTTAT 190251 AGGACTTTTA GGTTCTCTTA AGAGAAAATA GAGAAGAATA CTAAAAGCCA

190301 AGACTGCTAG AGCGGTAAGA ATAAATAATA GAGGAATGCC TAAAATTATA 190351 GCAGCGGTTC CAGAACTTAC AGCAATTCCT ATAATTAGAA TAAAAGCGAC 190401 AAGGTCCAGA CCTTTGAGCT TATGAGATGC GAGTTGCGTG GGCATGCTTT 190451 CGCTGCTGAT CAGCATGGTT AGATTTACGT TAGTCAAATT GGGTTCTGTG 190501 GTGGACATAA AACTTTTTAA ATAAAAAACA AAAAGTTTAA AAAAGATTCT 190551 TTTTTATGTG AAGTTATTTA TATTTTAAAT AGAAGTTGTT TATTTAAAAT 190601 AATAAATAGA CAACATTTTC ACTTTAAAAA GTATTGGATG CGGCTTAGAG 190651 CCAAGAATCT AGGGGGTGCT TGTGATCAGG ATAAATTGGC TGTGTTTTTA 190701 GAAAGCTGTT CGATTGGAAG CACGGACGCT TTCCGACATA GACTAGGGAG 190751 TCGTTAGATC GAAACGGGGT GGAATGATGG GAGGCTGGTC CTTGTCTGTA 190801 AGGATGATAA TTTTTTGCTC TTCCTGGTTG TCTTGTTCCC ATCCAAAATC 190851 TTGAAGAGAT GTGATGAGTG CCAATTTTAT AGCCTCGATC CATTCTGCTC 190901 TTGTTTTCCC GAGGTTTAGA AGTCTTGATG GAGCAAATAG ATTACATCCA 190951 ATGAGGGGA GTTGAATCAC ATCAACGCCT ATGATTTCAG CTTCTTGGAA 191001 CAGGTTATGA AACGCTTTCT TGCTTACTTC AAATGCTTGC TTAGGATCGT 191051 TGTTACACTT AGCAGCTTCG GGACCAAGCA GTTGTGCTAA GAAGTGTGCT 191101 TTGCCTGGGA CATGGTCGTT TGAGGTGTGA TCACTATTTT CCCATTTTGC 191151 TGAGCGGCAT TCTCCTGGCT GTAGTTGGGA TCCAGAACGA GAGTGCGCTC 191201 TAGGGAGCCT AGATGCGTTC CAACACTGTA GACTTGTAGC CAGGGATAGA 191251 GCTTTATTCG TTCCCCCTCC TTCTCGGGAG ATGTTCTCGT TTGCTGCGTT 191301 AACAATCATC ACCCTGCCTT GAGTTTTGAT TCTTGGAACT GCAATATCTC 191351 CTGAGGTAGA TATAAAGATA AGCTTCGAGT CTTTAAGCTT CCAAATAAAA 191401 CAGGGCTGCT GAGGTGTTTC TGTAGTAAAC GATGGGTCTA CAGCGGCTAA 191451 GCTCGGAAGA AGATCCCAGT TTTGTTCAAG AAGAGCCTCA TAGCCAGGAG 191501 TGACCGTGAT AGCGGTAAGT TCTTCTGGGG GTGTGGTTAG ATCAGCAAGA 191551 GAGACTTCGG GGAGTGATTC GGGATCGGGA AGTGGATCGA TCTGCAAAGG 191601 GAGGTCCTGT TGCTCAGGAG GCTGGGGAGA GGGAGACAGA GAACTTTGCT

191651 CAGATTCGGG TTCGATTTGA GGAAGAATCT CGTATAATTT AGGTTTCTTT 191701 AAAAGGGAGT AAAGTGAGAC CGCAGCAATT GCTAATGTTA TGACAACTAC 191751 GGCAGAAAA ATAGGCATAG CCAAAATGCC AGCAGCAATG CTAAAGCCCA 191801 AAGCGCACGC TAGGGCTAGT GTGAGGGTGA CAATTTTCAG TATTTTCTCG 191851 ATTCTATCTG TACGGTTGAG AGGGAGTCTA ATCGGATAGG AATGTTTCGC 191901 AGGAGTTCTG TAGAGACTGG CGTCTGTATA AGAGGGTAGG GGATTAGAAT 191951 CTGTCATAAT ATTTTATTAA TAGATCAATT TTCAACTATT ATATTACTAA 192001 TTTGTATTTT TATTAGATTT TTTTATAAAA CAAAATTAGT TTATTAATGC 192051 ATCCTATTGA AATAGATCTT TTTAGTTAAA AGGGACCATA AGTAGCTGTT 192101 TGTGGTCTGT AAGGATTATA GTCATGGGAG TTGAGGGGTG CTGCGCAGCA 192151 AAGGAGCGAA GAGCTTCCAT AAGACCTTTC TTTATATCAT TTACCCACTA 192201 CGTACGGATG TGGTAAAGCT TATATGCACT GCTATTAGGC TTTGTTTGGT 192251 TTACGGGTTC TAGTTCCAGC TTTCCTCCAG GTGAGTATAT AGAGGAAGAG 192301 ATCAGAGGCA CTTGGACCAC AGTGGCTTGG TTATTGAGAG CTTCATCAAA 192351 ACAGTTCAAA TAGGCTTTCT TAATAACATT GCTTAATTTC TCAGGATGTG 192401 CTTTCAATTC TCCTTCATAT TTAGGACCAA GAAGTTGTGC TAAGAAATGT 192451 GCTTCTCCTG GGTTCGTATC ATTTATTCGT CCAGTCTCTA TTGATCCAGG 192501 GTGCTGAGCG GCATTCACCC ACAGATAATC CTTTGCCAGT GTTTATTTTT 192551 CCCCCAGATG TTCTCGTATT GTTCCAACAA GTAGGGTGTG TGGCTGCTGA 192601 GAGAGCAGCA TTGGTTCCGG CTCCACCAGA TTGCATGTTC GAATTCGCTG 192651 CGTTAACAAT CATGACTCTT CCTGAGGTTT TCAGGCGTGG TTTAGCGATG 192701 TCTCCTGTAG TGGCTACTAA GGTAATTGGG GCTCCTTGAT GTTCCCAGAC 192751 ATAGTATCTT TGATTAGGAT CTTGGAGAGT CCAGGATATA TTGATTTCTG 192801 AGAGAGTATT GACTAGGGTC CATTCATTTC TCAGCAGCTT TTGGTAATCC 192851 AGAGGGACTG AGGGTTGCAC TTGAGCTGCA ATATCCTTGG GGACATGTTT 192901 TTGGGGCCCA AGGAGCTCTT CTGTCGGTTT GGATGGCTCT CTTCTTCTTA 192951 AAAGAAGACA AGAGAGTACG ACTGCTGCAA GGAGAGCAGC ACCAGTGGCT

193001 ATAGCCATGA GAGGCATACC AAGAACTCCA GCAACAACTG CAGTCCCTAC 193051 AGCTATGGCT AAAGCAAGAA TAAGAGCTGT GAGTTTTATT ATTTTAGCAA 193101 TCACTTCTTT CTTAGTAAGC ACGGGTGATT CTGAGATCAA TCTCTCATCT 193151 ACAGACAAAT GTGATTGATT CTCGTGTACA TCTGGGATTG AAGTGGTTTT 193201 AGTCATAAAA TATCTATTTA AAGGAAAAGA TTATATCTTA ATTTGTTGTT 193251 TTTTACAAAA AATCTTTTTC TTTGATAATA AAGATTTTTA TAAATTCTTT 193301 TTAATTTCAT TAATTTTAAT GAGTTAAGGC GTAGCAGGCG CACCTTATTG 193351 TTTAGTTCAT AAGAGAACTC TAAGTGGGTG CCACACAGAG TTTTTTCTCC 193401 ACTTGGTTCA AAAGGGTGGC GTTTGTGGTT TTATGGGGAA CGATTCTACC 193451 CCAGCCATGG GATTCAAAGA CTTCTTAGAT CGACAATACT ACTTGAGATT 193501 CTAATGGGGG GTGAAGACAT GTGTATTTCA TGGGCATGGG TAACGGAGGA 193551 CATCAGGTTT ACTTTCGAAG GCCCATATAT CATCTTCAGA GGTCAATAAA 193601 AGGGGGGGT AGGGAGAACG ATGTTTTACA ACAGTGTAGA AATTAACTAT 193651 TTGTGAAATA TTTTCATTTA ATAACAATTT AATGAGTTTT AATTTGAAGA 193701 CACAGGCAGT GGCTTATCTT TTGTCTATAA GTCGGTACAG TCAGACTTAT 193751 AAGAATTGCC GCTTTTTGGG TCTTCATAAT CTCTATCTTA AACAAAAAGG 193801 AGGTCAATCA CAAACTGGGT GAGTTGTAGA GGTCTCAATA AGAAAGTTTG 193851 TAAGGTTAGG AGAGCTATAA TTGCGCCTGG GATTCTGCAT AGACGATACC 193901 TGAAAAGATA GTGCCTAGGG TCGGTGTTTG CTGATGAGAA TGCCAAGAAG 193951 CATAGTCATC AACCGTCTTA ACATATTTAT AAGTTTTCTT AGCAGCGAAT 194001 GGCACAGCTG CTAGAAAGCA AAACAAGATG ATAACAGCCA CCAATAGAAT 194051 TGTTCCTGAG ACAGCAAGGA CTGCAATGGC AGCTGGATGT CCAGTTTCTA 194101 GAGTGAATTG CAGAGCAGGG CGGATTGCCA ATAAACTAAG AACTATGAGG 194151 CAGCATCCGA TCACTAGGAG AACTTCCAAT ATGACAGAAG TGATTGTCCG 194201 TGGAGAAGAC TCAGCTTGCT GCTTCCCCTT GAGTTCTATG AGTTTCTGAT 194251 TTTGTGTGTT TGCGCTATTG ATTAAGGGAG CGGCTGGACT TTCAAAAAGA 194301 ACGTCCGTAG CAGATACTGG AAGATATCCC ATTTATTTTA ATATCCTTTA

194351 ATTAAAATTT GCGCAGATAT TATACCTTAA TTTAATTTTT ATTAGAGAAT 194401 AAAATTCTAT ATTTTCTTTT GTGTAATATT TTTTTATAAA AACATGAGTT 194451 TTTAATTTAA CTTAGAAGAG AAGTGTTGCA ATATCTTTTA TTAAAACAAG 194501 AGTACCTAAG AATTCGTGCG TAAAATATTA CGTATTTGAT TTGTATAAGT 194551 CGTGAGAGTG TTTTTCTAT TATTGAGGGG CTATTTTGAT TTTTCTTGAA 194601 TTGCGTCATG GAAAGATCTT GGAGGCTTTA GGAGCCTTTT TGTCCCTCAT 194651 CATCTCGAGA TAACCAAACT GAAAAGATGC GTACCAAGAG TTTGTTTTTT 194701 CGTTCGCGAA TCTACTGTAC CCTTGATTTT AGGATAGCCG TTTTGAGTGA 194751 GATAGAGATC CTAAGAGAGA TCCTCAATAC TTTCTTGAGG TAGTTGGGCT 194801 AAGGATCTTT CAACACAAGC AATAGTTAGC GGATATTCTG GGTGCAGCTC 194851 GGCCATGTAT TTTACTGCTG TAATCATAGC CAAAGATTGT TGCATTCTTG 194901 ATTGCGGGGG CAATTCTCTA GAATAGAGAT GCTCCCTCAA GCTGTATAAA 194951 GGGATTTGGA TAATTTGGGT GTTCGTATTG ATCGCAGTCT GGATATAAGT 195001 GACATAGAGT GCATAGTAGT ATTGGTAGAT GGCTTCAGGA GTTGTTGGTA 195051 TTTGTGGGAG GAAGACGCGG CCTCGGAAAT CGGGTTGCCA GTATAAACGT 195101 TCTCCTGAAT CTCCCGCAGG GAGTTCTGTA TTTATTTCTA CCCTGGTTTC 195151 CTGAATTGCC TCATTGAGCA AGGATTTTGG GGAGGGAGTG GTATCTACGT 195201 AGGAGGATAT AGGCGCATTT TGCGAGGTTT CTTCAATTAA CAGAGTCTTT 195251 CCTAAATTTT CGGAAGCTGA AATTAAGAGA TGTTCTGGAT TGCCTCCCGG 195301 TAGGAGGGAT ACTAGGGAAA ACAGTGAATG TTTAGCTTTC CATATCTCGT 195351 AAGGGTTATC TGTGAGTAAT GCAGAGGGTG CGGGGAGATT CTGTAGGTCC 195401 TTGGGTTGGT GTTCCCAGTT TTCTTTAGAA CGATTGATTG CGATAGAGAT 195451 AGAAGAATAC TGTAGGTTGG TGGAGACCCT TTTTGGTTTT ATTTCTTTGT 195501 GTTTGCCACA AGCACGCGGT TTTAAGAGGG TGGTTGCTCG TTTATAAAGG 195551 ATAAAGTTGC TAAGAACAAC AGCAATAAAT GAAATTCCTG TGAGGAAATA 195601 GATAACAGGA ATATTTAAAA GCAAACCAAT AAGTAGGGTC CCTATAGTAA 195651 CTAAGACAGC AAAGATTGCA GTAATCAATA GAATTCGAGA CTGAATCTTA

195701 TCAAAGTTGC TGCAGGATTT TTTTTTAGTG TCTACTAGGG GATGATCTAC 195751 AGTTAAAGAT ATGGGTGAGA TTGTAGCCAT AATTAAGGAA TTCACCTGAG 195801 TTCCTTTAAA TTATATCTCT TTTATTGATA TTTAAATAAT TAAAAACATT 195851 TCTGCATTTA AAAATATTTT TTAATTTGTC ATCTTTCTAA GTAAATATAA 195901 AAAATAATAA GTTTATTTAG TTAAATTTTT TTTGAAAACA AGAAAGATAG 195951 GGGAAAGGCT CGCGAGTAAA TCAAGCCTTT CCTGAATGCA AAGAGTCGCA 196001 TAACTACTTA GAAAAGAGTT CGTCAGAATA GAAGTTGTAT TAATATAAAA 196051 TTTTTATTTT CTAAGATTAG AAAGTAACTT TGACACAGCC ACCTTTGATG 196101 CGGCACTGAC AAGCAAGACG TTCGTTAGAG TCTTCGGGTT CTCCTAGAAA 196151 ATCGTATTCT GGTTCCGTAA ACTCAGAAAG ATTCTCACGT CCTTCTAAGA 196201 CCTCTATCAC ACAAGTTCCA CAGACACCTT CTGTACAAGC AAAGGGAATG 196251 CCCATGGATT CACAAGGCTC TGCGATCTCA CTATTGTCTT CTAACTCGAA 196301 CTCTTGTTGT TCATCATCAG AGGTAATGAC TAGCTTGGCC ATGGAGTTTT 196351 CCTTCTACGT ATTGTCGGAT AGATTAAAAT GAAGTTCGGA GTAGAGGGAT 196401 TCGAACCCC GACCTATTGC TCCCAAAGCA ACCGCGCTAA CCAAGCTGCG 196451 CTATACTCCG TAAAATAAAA TTTTCGAGTA AAAGAGATCA ACGTTAGCAT 196501 AGAAGGAAAT AGTCGACAAG ATCAAAGATG CTTCATAACG TCTCTTTAGA 196551 GTTTTACTTG CAGATATATG GAAGAGGGAA GTATGAGAAA AGGCAGGGAT 196601 TCTCTAGGAG GCTGTTTTTG TGTCTGGGAA GAAAGATGGT GTAAGGGGAA 196651 TGATCTTTGT CCCTCTTAGC ATCCTAGTAC TAATCTTTTT ACCTCTTCCT 196701 CAGATCCTTC TTGATTTTGG ATTGTGTATT AGTTTTGCAT TGTCTTTACT 196751 AACGGTCTGT TGGGTCTTTA CCTTAAATTC AAGCAATTCA GCGAAGCTTT 196801 TTCCTCCATT TTTCTTATAT CTTTGCCTAT TGCGGTTGGG ATTGAATCTT 196851 GCATCAACAC GATGGATTGT CTCTTCAGGA ACCGCCTCTT CTCTGATTGT 196901 TTCTTTAGGC AGTTTCTTCT CTTTAGGAAG TCTATGGGCA GCAACGTTTG 196951 CGTGCCTCCT TCTTTCTTT GTGAACTTTT TGATGGTTTC AAAGGGTTCG 197001 GAAAGAATCG CAGAGGTCCG TTCGCGGTTT TTCTTAGAGG CTCTTCCAGC

197051 AAAACAGATG GCTTTAGATT CTGATCTTGT TTCTGGAAGA GCTTCTTATA 197101 AGGCTGTCAA AAAACAAAAA AATGCCCTTA TAGAAGAAGG GGATTTCTTC 197151 TCTGCCATGG AGGGGGTCTT TCGTTTTGTT AAAGGGGGATG CAATTATTAG 197201 TTGTATCCTT TTACTCGTGA ACGTAGTTTC TGTAACTTGT CTTTATTATA 197251 CTTCGGGTTA TGCTCTTGAG CAGATGTGGT TTACAGTTTT AGGAGATGCT 197301 TTAGTGAGTC AAGTACCTGC TTTACTTACT TCGTGTGCTG CAGCCACTCT 197351 TATTAGTAAA ATCGATAAGG AAGAGAGCCT TTTAAATTAC CTGTTCGAAT 197401 ACTACAAACA GTTGCGTCAG CATTTCAGGG TGGTGTCGTT ATTGATCTTT 197451 TCTTTGTGCT GCATTCCCAG TTCTCCAAAA TTCCCTATCG TTTTGCTCGC 197501 GAGTCTTTTA TGGTTGGCGT ATCGAAAAGA AGAGCCTGCA TCAGAAGATT 197551 CTTGTATAGA ACGTGCGTTC TCTTATGTTG AGGGGGCCTG CCCTAAGGAA 197601 CAAGAATCAC AGTTCTATCA AGTATATCGT GCAGCATCCG AAGAAGTATT 197651 TGAAGATTTA GGAGTTAGAT TGCCTGTGCT TACTTCTCTA CGTATTGAAG 197701 AGCGTCCTTG GCTCCGAGTA TTTGGCCAGA ATGTATACTT AGATGAAATG 197751 ACTCCAGAGG CTGTGCTTCC TTTCCTTAGA AACATCGCTC ATGAGGCTCT 197801 CAATGCCGAG GTAGTTCAAA AGTACCTTGA GGAATCAGAG AGAGTGTTTG 197851 GCATCGCTGT TGAAGACATC GTTCCTAAGA AAATCTCTTT AAGCTCTCTT 197901 GTAGTTCTTT CTCGCCTCCT TGTTAGAGAA AGGGTATCGC TTAAGCTTTT 197951 CCCAAAGATT CTAGAGGCCG TTGCGGTATA CCAAAATTCT GGAGACAGCT 198001 TGGAGATCCT TGCGGAAAAA GTGCGAAAGT CTCTCGGATA TTGGATTGGG 198051 AGAAGTCTCT GGGATCAGAA ACAAACCCTT GAGGTAATTA CCATAGATTT 198101 TCATGTTGAA GAATTGATAA ACAGCTCATA CTCAAAGTCT AATCCTGTAA 198151 TGCAAGAGAA TGTGATCCGT CGAGTAGACA GTCTTTTAGA ACGGTCGGTA 198201 TTTAAAGATT TTCGAGCCAT AGTTACGAGC TGTGAAACAC GATTTGAGAT 198251 GAAAAAATG CTCGACCCAC ATTTCCCTGA TCTTTTGGTT TTATCTCATG 198301 ATGAGCTTCC TAAAGAAATC CCTATTTCCT TCTTAGGGAT CGTTTCAGAT 198351 GAGGTTTTAG TTCCTTAATT TAATTTAATC TTCTGTAAGC ACTATTACTT

198401 TCGTATTTT TTTGTTGGTG TAAATATTGT TTAAAAATTT TTTTTGAATT 198451 AATCTAATAA ATAACTAGAT AAAAAAAAT TTGTGAAAAC ACAGCAAACT 198501 CAAAACATCA TAGAGGTTTG GAACTTCTAC TGGGAGACTC AGGAAATAGA 198551 GTATCGCGAT AGCTTAATTG AGTTCTATTT GCCTTTAGTA AAAAGTGTGG 198601 TTCATCGTTT GATTTCAGGG ATGCCTTCCC ATGTAAAGAC CGAGGATTTG 198651 TATGCTTCGG GTGTTGAAGG TCTCGTCCGT GCGGTGGAAC GTTATAATCC 198701 TGAGAGAGT CGTCGTTTTG AAGGTTATGC GGTATTTCTG ATTAAGGCTG 198751 CCATTATTGA TGATCTGCGT AAGCAAGACT GGGTTCCTCG TAGTGTCCAT 198801 CAAAAAGCGA ATAAATTGTC AGGAGCTATG GATTCTCTTC GCCAGTCTTT 198851 AGGCAAGGAA CCCACGGATC TTGAACTGTG TGAGTATCTC AATATTTCGC 198901 AACAAGAGCT TTCGGGATGG TTTGTATCTG CCCGTCCTGC ATTAATCGTG 198951 TCTCTGAATG AAGAGTGGCC TTCACAAAGT GATGAAGGAG CCGGAATGGC 199001 TCTTGAAGAG AGAATCCCCG ATGAACGTGC CGAGACAGGG TACGATGTTG 199051 TAGATAAACA AGAATTTTCT TTATGTTTAG CCAATGCGAT TCAGGAACTT 199101 GAGGAAAAGG AACGCAAGGT CATGGCCCTG TACTACTATG AAGAACTTGT 199151 CCTTAAGGAA ATCGGTAAGG TCCTTGGGGT AAGTGAGTCT CGCGTCTCTC 199201 AAATTCACTC TAAAGCATTG CTTAAGCTTC GCGCAGCACT CTCTGCATTT 199251 CGATAAATAC AGTTCTCAGG TTTTAAGAGC AGTCCTAGAG CTAGGAGAAG 199301 CCCTCCTACG ACACAGAGTA ATCCGTAAAG AATTTGTTTA GTCAACGCCA 199351 TAACAATGCA GCCAGCAATT GTGAAGATTG CTCCTACAAT CAAGAGAAGA 199401 ATGCGGGGCA GCCATTCTTT AATACGTTCG GCTAGGGTAG AGGTAGGAAG 199451 GGGTGCCTTG CTAGTTGTGG CATTTAGACT AGTTGCGGCG TTTAGCATCG 199501 ACGGAACCCC AATTGGAGAT GTAGACATAG CGACCTAATC TTTAGAGAAG 199551 ATAGAAGTTA TGGTATCTCA GCAAAGGAAT TATTCTCAGC GTACAATCTT 199601 TTCTTTATTG TTAGATAGGT GGTGCCTTGC ATATAAGCTT TAGGATCGAT 199651 AAGATAGAGC TTGGCAATTC TTAATTATGT ACGATCACTC ATGCAATCCT 199701 GGTTACAATC TTTACAAGAG CGAAATATTT TAGAGAATTT TACCGCAGGT

199751 TTGGAATCCG TAGAGGGACC TATCGCCGCT TATTTAGGAT TTGATCCTAC 199801 CGCACCTGCT CTACATATTG GTCATTGGAT TGGGATTTGT TTCTTGAAGA 199851 GACTCGCTGC TCTGGGGATT ACCCCCATAG CTTTAGTCGG GGGAGCCACA 199901 GGTATGGTTG GAGATCCCTC AGGGAAACAG AGCGAGAGAT CGTTACTTCA 199951 GACAAGTGAA GTTTTTGATA ACAGTCAAAA GATCACGGCG TGTCTCCAGC 200001 GCTATCTTCC CGGGGTGACT CTTGTAAATA ATGCAGACTG GTTGCAGGAG 200051 ATCTCCCTGA TTGATTTCTT AGGGGATATA GGAAAACACT TTCGTTTAGG 200101 CCAAATGCTA GTGAAAGATA CAATAAAGCA GCGGGTGCAT TCTGATGAAG 200151 GAATTAGCTA TACCGAGTTT AGCTATTTAA TCCTGCAATC CTATGATTTT 200201 TATCACTTAT TTAAAAATTA TGGCACGATC TTGCAGTGCG GTGGTAGCGA 200251 TCAGTGGGG AATATTACTT CAGGAATCGA TTTTATTCGC CGTAAAGGGT 200301 TGGGTCAGGC CTACGGCCTT ACCTATCCTT TATTAACGAA TGCTCAGGGG 200351 AAAAAATAG GGAAAACAGA GTCGGGAACT GTATGGCTCG ATTCAGATTT 200401 AACCTCTCT TTTGAGCTGT ACCAATACTT ACTCCGTTTG CCCGATGATA 200451 CCATCCCTAA AATTGCTCGT ACGTTAACTT TATTGAGCAA TGAAGAAATT 200501 CAAGATATTG ATAGGCGTGT ACAGACGGAT CCAGTTGCAG TGAAGGAATT 200551 TGTAGCCCAA GATATCTTAA GTGCTATTCA TGGAGATCTA GGGCTTGAAG 200601 AGGCTCTTTC TGTAACTCGT AGCATGCATC CAGGGAATCT TTCATCCTTA 200651 TCGGAAAAAG ATTTTCATGA ATTGTTTGCA GGAGGGATGG GGGCCTCATT 200701 GGATAAATCC GAGGTGTTAG GGAAACGTTG GTTAGACCTA TTTCTTGTTT 200751 TGGGACTATG TAAATCTAAA GGGGAAATTC GAAGGCTAAT TGAACAAAAA 200801 GGGGTATATA TTAATAATGT GCCCATCGCT AATGAGCATA GTGTTTGTGA 200851 AGAACAAGAC ATCTGTTATG GTCACTATGT GTTGTTGGCT CAAGGTAAAA 200901 AACGAAAGCT TGTTCTATAT TTAAATTAGT TTAAGGAGGC TAGGTAGCTT 200951 TGCAAACGAA TATTGGTCTT ATTGGCTTAG CTGTCATGGG GAAAAATCTT 201001 GTCTTAAACA TGATAGATCA TGGTTTTTCT GTCTCTGTCT ATAATCGGAC 201051 CCCAGAGAAA ACGCGGGACT TCTTGAAAGA ATACCCTAAC CACCGAGAGC

201101 TTGTAGGGTT TGAATCTTTA GAAGATTTTG TGAATTCATT GGAGAGACCA 201151 CGAAAGATCA TGTTGATGAT TCAAGCAGGG AAACCTGTGG ATCAGAGCAT 201201 TCATGCGTTA CTGCCTTTTC TAGAACCCGG CGATGTGATT ATCGATGGGG 201251 GGAATAGCTA TTTTAAAGAT TCCGAACGAC GATGTAAAGA GTTGCAAGAA 201301 AAGGGGATTC TCTTCTTAGG CGTGGGGATT TCTGGAGGAG AAGAAGGTGC 201351 ACGTCACGGC CCATCAATTA TGCCTGGAGG AAATCCTGAG GCGTGGCCAT 201401 TAGTGGCTCC TATTTTTCAA TCAATAGCAG CAAAAGTACA GGGCCGTCCC 201451 TGCTGTTCTT GGGTAGGAAC TGGCGGTGCA GGCCACTATG TAAAGGCTGT 201501 TCACAATGGT ATAGAATACG GCGATATCCA GTTGATATGC GAAGCTTACG 201551 GTATCTTAAG AGATTTCCTA AAGCTCTCCG CAACTGCCGT TGCTACAATT 201601 TTGAAAGAGT GGAATACTCT AGAGTTGGAA AGCTATCTAA TTCGTATTGC 201651 TTCTGAAGTC CTAGCATTGA AAGATCCGGA AGGAATCCCT GTTATTGATA 201701 CGATTTTAGA TGTCGTGGGC CAAAAAGGTA CAGGAAAGTG GACCGCAATC 201751 GATGCTTTAA ATTCTGGAGT TCCCCTTTCC TTAATCATAG GAGCTGTTCT 201801 TGCTCGTTTC CTTTCTTT GGAAAGAGAT ACGCGAGCAA GCTGCCCGTA 201851 ATTATCCAGG AACCCCCTTA ATATTTGAAA TGCCCCATGA TCCCTCGGTA 201901 TTCATACAAG ATGTCTTTCA TGCTTTATAC GCTTCCAAGA TCATCAGCTA 201951 TGCTCAGGGA TTCATGCTTT TAGGAGAAGC TTCAAAAGAA TATAATTGGG 202001 GATTAGACCT AGGAGAAATT GCTTTGATGT GGCGCGGGGG ATGCATTATT 202051 CAAAGTGCAT TTTTAGATGT TATACATAAA GGATTTGCTG CCAACCCAGA 202101 GAATACCTCG CTCATCTTCC AAGAATATTT CCGTGGAGCA TTACGCCATG 202151 CGGAGATGGG ATGGCGTAGA ACAGTAGTGA CTGCAATTGG TGCAGGGCTA 202201 CCTATTCCCT GTTTAGCAGC AGCAATCAGG TTTTATGATG GCTATCGTAC 202251 AGCAAGCTCT TCAATGTCGT TAGCTCAAGG ACTGCGAGAT TATTTTGGAG 202301 CTCATACCTA CGAGCGTAAC GATCGCCCTC GAGGAGAGTT CTATCATACC 202351 GATTGGGTGC ACACGAAAAC TACAGAAAGA GTGAAGTAAA AATAAAAAAT 202401 CTCGAGAAGA CCCTAGGTAG CTCGAGATTT AGTTCACCAC TACCGAATTT

202451 TCAATTGTAA GCAGTGTGCT GATTTTAAGC GTCAATGTTA ATCTAATTTT 202501 AGAACTTCAA TGAAAGCTGT ATTGGGAATG GAAACTTTTC CAAATTCCTT 202551 CATACGTTTT TTTCCTTTCT TTTGCTTTTC CCACAGCTTG CGTTTCCTAG 202601 TAATATCTCC GCCATAACAC TTTGCGGTCA CGTTCTTAGA AAGCGCACGA 202651 ATCGTTTCTC TGGCAATGAC TTTTTTGTTA ATGGCAGCTT GGATGGGAAT 202701 CTTGAAGAGT TGTTGTGGAA TCACGTCCAC AAGCTTTTCG CAGATACTTC 202751 TTCCACGAGA TTCTGCTTTA TCTCTATGGA CTAAACAAGA AAAAGCATCT 202801 ATGGGCTCCT CGTTAATAAG AACCTCTAAT TTGATGATCG ATCCCTTACG 202851 GTAATCCCCA AGACGGTAGT CAAAGGATCC ATAACCTTTA GTTACTGACT 202901 TCAGCTTGTC ATTGAAATCC GAGACAATCT CATTTAAAGG GAGTTCGTAA 202951 GCAAGAACTA GACGGTGCTG ATCTAGCATT TCTGTTTTTA CGCAGATCCC 203001 ACGTTTATCT AAACAGAGGT TCATAATGTT GCTCAGATAT TCTTGAGGGG 203051 TGATAATATT CACATGAACC CAAGGCTCTT CCACATGCTC GATGATCGCA 203101 GGATCCGGAT ATCCTGAGGG GTTATCAATA TCTAGAACTT TCCCGTTTTT 203151 TAAGACGACT TTATAGATGA CACTTGGAGC CGTTGCAATA ATATCTAAGT 203201 CAAATTCTCG AATGATTCTT TCAAAGATAA TCTCAAGATG AAGAAGTCCT 203251 AAGAAGCCAC AACGAAAACC AAAGCCTAAA GAGTGACTGC TTTCTTGTTC 203301 TATAGTTAAA GCAGAATCAT TGAGCTGTAG TCTTCCTAAA GCATCTTTCA 203351 AAGTATCAAA ATCAGAAGAA TCTATAGGAT AAATTCCAGC AAAAACTACC 203401 GGATTGATCT CTTTGAAGCC TTCCAAAGGA GTTTTTGCAG GATGTTTTGT 203451 TTTCGTGACT GTATCGCCGA TCTTCACATC CTTCACTTTT TTGAGATTGG 203501 CAATAAAAA ACCCACCTGA CCAGGGCGTA AGGAACCTTC TATAAATGTT 203551 GCTTTAGGGA GAAAGGCCCC TATACCTAAG ACTTCAAACG AGGAGCCTTT 203601 AGCCGCCATA AAAGTAATGC GGTCTCCTTT TTTTAATTCC CCGCTAATAA 203651 TGCGTACGTA GACCATAATG CCAACGTAAG GGTCATAATG AGAATCAAAG 203701 ACTAAAGCTT TAAGCTCTGT TTCTGCAGGT GCTTTTGGAG GAGGAACAAG 203751 ATCGATAATT GCTTTCAGGA TTGCAGGGAT CCCCTGACCT GTTTTTGCAG

203801 AACAGGCAAT AATGTTCGTA GTGTCTAGGC CTATATAATC TTCAATCTGT 203851 TGAGCAATTC TCACGGGATC AGCGGCAGGT AGATCAATCT TGTTTAATAC 203901 AGGAATGATC TCTAAATCTC TTTCAAGGGC CAGGTAGACA TTAGCAAGAC 203951 TTTGTGCCTG CACCCCCTGG GCGGCATCTA CAATAAGTAA GGCGCCCTCA 204001 CATGCAGATA GAGATCGAGA GACTTCATAC GAAAAGTCCA CGTGACCAGG 204051 GGTATCAATC AGGTTCAGTT GATACACCTC TCCTTCATAT AGATACGTCA 204101 TGGTGACAGG ATGAGCTTTA ATTGTAATGC CACGCTCTCT TTCAAGATCC 204151 ATGGAATCTA AGAGCTGCTC ACGCATCTCC CGTTCTTCTA CTGTGCTCGT 204201 ACTTTCTAAA AGGCGATCAG CAATTGTAGA CTTCCCGTGA TCAATATGCG 204251 CTATGATTGA AAAATTGCGA ATGTTCTCTA TCTTATATTC TTTCAAAATG 204301 TACTGTAGTG TTATCTAGGT TTATTCCTGG TTTCATAAAG CTGCATTGAG 204351 AAGGCTCCTT CAGACAAAGA CCAGCTTCTC AAATACTTTT GCGGACTGCA 204401 TCTCTTTATT AGCAAAATGA TCATGCGGTC ATCTTTAGCT TTGGATAGAT 204451 GAAAGTATAA TAGATGAGAT GTAGAAACCA CAAGGGTTTA AAGTCGAATC 204501 TGATTTGAGG TAGAGTTCCA GAGTGGCTAG CAGTAGAAAA AACTAAGTGA 204551 GAAGAAGTGC TCTCCGTTAG TATGAAACTG ATTCCCACTC AGGATTCTAT 204601 AGAAAGGGAA ACCGATTCTA AAAGAGATAA AAAAATATTT ACCATTTACA 204651 TATGTTCATC TAAAGTCCTT GCGGGTCATT TTTTCAGTCA TTTAGACAAG 204701 CATAATAAAA TTCATGAAAG CATTGGGGTT TGAGATAGTT AAATCGACTC 204751 GATCCAAGAG TATAAGAGAG GAATGAGTGT TCTTATGTCC GAGATAGTGC 204801 TCTTCAATGT AGTGAAGAAC TGGAAAGTCA GGAGATCTAT TTGAGTAGGA 204851 GTTCAGGTCT TCGCAATCGA TTTTTTCAAG CTCTGGAGTC TATGTGAAAG 204901 AAATTTATAG AAGTAAAATA ATTCCATGAA TTCTAAAATG ATTAGGAAAA 204951 TAATATAACA CGCTGATACT CAAGTCACAA TGGAGCTTCT ATGGTTAATA 205001 TACAGCCTGT GTATAGGAAT ACCCAAGTCA ACTATAGTCA GGCTACCCAA 205051 TTTTCGGTGT GCCAGCCAGC GCTTAGCCTG ATTATCGTTT CTGTTGTTGC 205101 TGCTGTACTC GCTATTGTAG CTTTGGTATG CAGTCAATCT CTTTTATCCA

205151 TAGAGTTAGG AACTGCTCTT GTTCTAGTTT CTCTTATTCT TTTTGCTTCT 205201 GCTATGTTTA TGATTTATAA GATGAGACAA GAACCTAAGG AGTTGCTGAT 205251 CCCTAAGAAA ATCATGGAAC TCATCCAAGA ACATTATCCA AGTATTGTTG 205301 TTGATTTTAT TAGAGATCAG GAGGTTTCCA TTTATGAGAT ACATCACTTG 205351 ATCTCTATTC TTAATAAGAC GAATGTTTTC GACAAAGCAC CAGTATATTT 205401 ACAAGAAAA CTCTTACAGT TTGGCATTGA GAAGTTCAAA GATGTACATC 205451 CAAGTAAGCT CCCTAATTTT GAAGAAATTC TTCTACAGCA TTGCCCATTG 205501 CATTGGTTGG GACGTCTGGT ATATCCCATG GTATCGGATG TCACTCCAGG 205551 AACCTATGGA TACTATTGGT GTGGTCCTTT AGGACTGTAC GAGAACGCTC 205601 CCTCTCTTT TGAACGTCGA TCTCTTCTAT TGTTAAAGAA AATTAGCTTT 205651 GGAGAGTTTG CTCTTTTAGA AGATGGTCTC AAGAAAAACA CGTGGAGTTC 205701 TTCGGAACTC GTTCAAATCA GACAAAACCT TTTTACAAGA TATTATGCTG 205751 ATAAAGAAGA GGTAGATGAA GCAGAGTTAA ACGCTGATTA CGAACAGTTT 205801 GATTCCCTCC TTCACCTTAT TTTTTCTCAC AAGCTCTCTT GAAAGCAAGT 205851 GCAATTATTT CAATATATGA ATGAGTCCGG ATGGGATTGG CTTTGTGATT 205901 TTGATTCTCA AGGCGAGGGA TTCCAGTTAT CACGTCTGGT TGGGCTGTTA 205951 CATTCGTCCT GGGCATTATA CGAAGCAAAA GAGCAATTTT ACCTTCCTGA 206001 GGTTTCTCTA TTGACCTGGG AAGAACTGAT AGAAATGCAG TTATTAAGCA 206051 AACCAACAA ACACGGGGTT GCAAAAGATC TTTGTAATGT ATTTGAAAAA 206101 CACTTCAAA GGTTTAGACA GTACCTAGGT TCCTTAGATC TAAATCAAAG 206151 GTTCGAAAAT ACCTTCTTGA ATTATCCTAA ATACCATTTA GATAGGGAGT 206201 GAGAAAAAA TCCTAGGTCA GCTTGGCAGA AATTTTTGAA ATCCTTAAGT 206251 GTTCGATCTG CATTTTTTC GGGGATTGTA AAAAGTTTGC CTAATCTGAA 206301 GGGTAAGAGA GAGCTATTTT TCATGGGATT TTTTATTACA GGAAATTCTT 206351 GAATTAGAGA TTTTTATTTA CACATAATCT ATACTGCCTT CAATAGATCT 206401 ATATCTAAGG AGTTGGCTAT GAGCATGACG ATCGTTCCAC ATGCTTTATT 206451 TAAAAATCAT TGCGAGTGTC ATTCTACCTT TCCTTTGAGT TCAAGGACTA

206501 TTGTAAGAAT AGCCATTGCC AGCCTCTTTT GTATAGGTGC ATTAGCAGCT 206551 TTAGGCTGTT TGGCTCCTCC CGTTTCTTAT ATTGTTGGGA GTGTTTTAGC 206601 TTTTATTGCC TTTGTCATTC TTTCTTTAGT AATTTTAGCT TTGATTTTTG 206651 GAGAGAAGAA GCTTCCACCA ACACCAAGAA TCATTCCTGA TAGATTTACT 206701 CACGTGATAG ATGAAGCTTA TGGCCTTTCA ATCTCTGCAT TTGTAAGAGA 206751 ACAGCAGGTA ACATTAGCCG AGTTTAGACA ATTTTCTACT GCCCTGTTGT 206801 GTAACATATC TCCTGAAGAG AAAATCAAAC AATTGCCTTC TGAATTGCGA 206851 AGTAAAGTAG AGAGTTTTGG TATTAGCAGG CTCGCAGGTG ATTTAGAAAA 206901 GAATAATTGG CCAATATTTG AAGATCTTTT AAGCCAAACC TGCCCGTTAT 206951 ATTGGCTTCA GAAATTTATA TCAGCAGGAG ATCCACAAGT TTGTAGAGAC 207001 CTAGGTGTCC CTAGAGAATG TTATGGGTAC TATTGGCTAG GGCCTTTGGG 207051 ATACAGTACA GCTAAGGCTA CAATTTTTTG TAAAGAGACG CATCATATTC 207101 TTCAACAATT AACGAAAGAG GACGTTCTTT TATTAAAAAA CAAGGCTCTT 207151 CAAGAGAAAT GGGATACTGA TGAAGTCAAA GCAATTGTAG AGCGTATCTA 207201 CACTACCTAT ACGGCACGAG GAACTCTAAA GACCGAAGCA GGGGGACTTA 207251 CAAAAGAGAC AATCAGTAAG GAATTGCTAT TGTTGAGCTT GCATGGCTAT 207301 TCTTTTGATC AGCTACAGCT GATCACTCAA CTTCCTAGAG ATGCTTGGGA 207351 TTGGCTGTGT TTTGTAGATA ACAGTACCGC ATACAACCTT CAGCTTTGTG 207401 CTCTTGTAGG AGCTTTGTCA TCCCAAAATC TTCTTGACGA ATCTTCTATC 207451 GATTTTGATG TAAACCTAGG CCTGTATGTG ATTCAGGATC TAAAAGAAGC 207501 TGTTCAAGCA TTTTCTGCTT CTGATGAGCC AAAGAAAGAA CTAGGTAAAT 207551 TCTTGTTAAG GCATTTGAGT TCAGTTTCTA AGCGATTAGA GAGTGTATTA 207601 AGACAGGGTC TTCACAGAAT AGCTCTAGAG CATGGAAATG CCAGAGCTAG 207651 GGTTTATGAC GTCAATTTTG TAACAGGAGC TAGAATTCAT AGGAAGACGA 207701 GTATCTTCTT TAAAGACTAA ACCAGGTAAC TAGCTTTTTA GTCTCGAAAG 207751 GGGCTACGAG AGGGAGTAGA ATTTTTCTTA TTCCAAAATA GAAATCTCTA 207801 ACCTATCAGC AGGGAGGACT CGTCTTCTTC AGGTTGGGAG AAGACGCTTC

207851 TCTTGAGTGG ATGGACTGTT ATAGGTGGTA AGCTTCTAAA GGTGGAATTA 207901 TGAGTTGCTA AATTCTCAAG GAACATCGCT ATATTGGTTG ACTGACGACG 207951 GATCGTTTTA AACTGATCTA AACTAAGGAG TGTCAATGAA GAAAGAAATG 208001 CTTCGCTTTC TTTATGAATT AGGCCTTGTG TGTAGAGAGT ACCAATTAGA 208051 GAGGCAAATT CCATGTCTTG GGGTCGTCCT GTATGGTCTA AAGTCAAGCA 208101 ACAGAGGTTT TCCCATAAAT CAGCCGGGAC TGTTTTTATT AAGGACATCT 208151 GCTTCCACGA GTACCTATGC TTAAAAATAC ACAAGAGCAG GTTATAGAAT 208201 TCATCCTCTT GAAGAGAGGG TTGAGATAGA TGTTTATAGG ATTGATACGC 208251 CGCATGATAT TTAGCAAAGA GCTGCTGTTT CATAGTATTT AGATCACTAT 208301 GATTCAGGTT CCACTCGCCA TTTAATGCTG CATACTTAAG ATGTTGAAAA 208351 ACATCACGAG TTAAGACTCT AGCTAAAACT AATGTGTGTA GATTTAGAAT 208401 AGAGGAATTT GTTGGTCTAA AATCTAAAGG AATCAACCAA TAGGTAGGAA 208451 GCATCTCTGG TTTTATATAA TCACTAAGGT CTTCTTCGAG GCTGCCTCCA 208501 ATTTTCTTAA CTCGATGGAG CTCTACAACC TTGCTGCCTG CAGAAATAAA 208551 CTCTCCTAAC CAGAAAAATG GAAATACCTT TTGTAAGATT TTTGGTAAAG 208601 AGGGGAAGAG GGGACGGGAA GAAGCAGCGC TTTCAAGTCT TTTAAGAGAA 208651 GCAGTACGGA CTTCATTTT TAAGCGTGCG ATACCCTCGA ACGTATCTAT 208701 TTTCTGTTGT AGTTCTTCAG ATACGTTGTA ATTTGTAGAT GATCCAACTT 208751 CAGAGTGCAA TTGATTTAGA AGATCAATAA AACTTATGAG ATCTTTAAGA 208801 TTTGGTTTAG CTTCTGAAAC AAAATCAGAG ACAAATTTAG GATAGTGCGC 208851 CCTGATTCTA TTTACGAGTT CTTGGGGGAA TACTTGTTCT TTTGATGAAA 208901 TCATAGGCGT GATTTTTTT ATTCCTAAAA TCACACCAAT CAAGGCTATT 208951 AATAATCCTA ATCCTAACAA TGCGCCACTT AGAATATAGG AAACAGGAGC 209001 TGCTACACAC AAGAAAGCTA TCAAAGCTCC GCAGAGTAAG ATGGCACTGA 209051 TAACAATATG AATAGTGGTT GAATTCTTTA ATTCAAAATA ATAATTACAA 209101 GAGCGATTAT TTTGAATAAC TGGCGAGGTT ATATTACTCA TACATTTCCT 209151 AGGCTCATCA ATTTGCGTAT TATATCGTAA ATTTAACAAA AATCCTATTA

209201 ATAGAAATGT TTTTATTTTT AAAATTTTTA TTGAAATTGT TTTGTATTAT 209251 TGATGAAAAG TTTTCATTGA AGATAGCTTC TAGAAGAAGA ATAACGAGAG 209301 AGCTTGTTAT GCTTTTTTTA AAAAAAGGGC GAGCCCCTTT TTGTTCTTTT 209351 CCCGATTGGG AATAACGTAT TCTTTTTATA AAAATTAAAG ATTTTTTTAT 209401 ATTTCTAAAT CTCAATTATA AAAAGTCTTT CAAACGCAAT TTGCTTGTAG 209451 GTATTGCAAA TGATCAGACT CCCCTTATAG GGAAGATCGT ATAGGGAGAC 209501 TGCTTTCTAT TTTTTCTCT AGTTTGTAGG TTTGGGCGTG GAAAGCATAC 209551 TGAGGTAAGC TTGATGTCCA TTGTGAACTC GTATTACGAA CAAATTGCCA 209601 TCGTTTTAAC AGATTTTGAT GGTGCGTTGT ATTTACGCAT ATCTTCTGAA 209651 CAAGAGCACT CGCTGGGTGC ATAGGACTTT CTTTTACTTT CTCTAGTAAA 209701 ACCTTCAATT CTTTCCACGT CATGAAGTTC ACTGTAGGTT CATAGTTAGA 209751 GGATACTGGA TCGAACATAT TTGTTTCAGT ATTCAAAAAG CCTCCAAATG 209801 TTGCCATGGA ACAGTGGCCT CCTGCTTTAT CAAACTGACA CAACATTTTC 209851 CAATTATCAG GATTTATAAG TTGAATCATC TGAGCCTGTT CCCAAGTGAT 209901 ACCATGAGAA AAGAAAAGAA ATAAGAATTG TGAGATTCCT TGAACATCCT 209951 TCCGGAAAAT CATATTGTGG GGGAGTTCCT TGAATATTTC TTCGCAGGTT 210001 TTTTTCACAG AAGGAGAATC CCATTGATTC TTAGACGCTT TACTATATAG 210051 GAACTTATAC TGTGATTCTG AGATTAATGT TAGTAGAGGG CGTGTATAAG 210101 AGTGGAAAAT AGTTGTATAT CCTTTATGAA ACGCTAAGGG CCCAAGTAAA 210151 CCATAAACTT TTTGTGTTTT ATTTAATCCG ATTTCCCCAG CAACAGATTC 210201 AGTTTTGTCT ATAAAATGGG AGAGCCAGTA TAACGGGCAG TTTTGAAGAA 210251 GAATCTCTTC GAACTCTGGA AACAGGGTTA AATCTATAGA TTTTAGAATA 210301 TCGATCCGA AAGCCTCTGC TTTTTTATGT AAATTCGGAG GCAGGTCTGT 210351 ACCGCTTTTC CAGCAATTAA TAAATATTTT TAATTCGTTA ACAGTCAGGG 210401 AGTGTGTTTT CACGAAATAA AAGACTTCTT TAGGATAGCG ATTGTAAATA 210451 ATCTTCTGAA GTTCGTTGGG GATGGGCAGA ACCTTTGATT TAGCTAGCAG 210501 TGCTACGACT AGCGTTATAA TCAAGATTAC GATAGCGGCT AAAGCTAAAG

210551 TTCCTCCAAT AGCATAGCTA ATCGGCGCAG CAAGGAAAAC AAAAGAAAGT 210601 GCGCTAACAA GAGCTAGAAC AAGCCCAAGA ATGAGTCGGG CAATTGTCCT 210651 AATTTTTAAA GAACAAGATC TATGACATTC GCAGTCATTT TTTAAAAAATA 210701 TATGGGGAAC TAGTGTTAAA GGACTACACT TCATAGTACA ACCAGAATGT 210751 CTCACGGTTA TTATAACCTT TTGATTGAAA AAAGCGTACG AATTCGTGAA 210801 ATGGGAGTTA ATAAAAAAA TCCCTACCCA CTAAAAAGCA GGTAGGGATC 210851 AACAAGAGTA AGAGAAGCAA CTCTATGAAG AAGCAGGAGC TGAATCTTCT 210901 TGAGCCACTT CTTGTTCTTT AAGAGCAGAC TGCGCTAAGA ATAGTTTGTT 210951 TAACTTAGTT GCAGAAACCA ACCAAATAGC AATGATGAAA AGAAGAATCA 211001 CTGCAAGATA AGGGGTCATA GCTCCAATAC TTCCACAGAT AACGAGCAAA 211051 CCTTGTTGGA TTAAAGCTCC TCCTGATTTT CCGAAGCGGG CAGCAACTAC 211101 ATCAATAGCA GCCTTACCTT TGACTTTTTG CTCTTGGTCA AGAGGGATAT 211151 AGGCCATTTC TTTAGTTGAG TCAAAGAGAG CGTATTTTGT GGATTTCGAA 211201 AGAATATTCT GTATAGCTCC GACAACCACA GCTAGCATGA GAGGAGTTGT 211251 ACCGAACATA GCGACCAGCC CAGAAGCTTG GTTTCTAAAG ATAACAAGAG 211301 CGAAGAAAC GATACCTGTT AGGAGAACCA TGACAGGAGT GACTAGGGCT 211351 CCAGTTAACC ATCCAAATTT ACGAATGACG TTACCACCAA CAAATAGCAT 211401 GATAAGTACG GATACTACGC CAGTCCAGAA GGAGAAGTTC CCCATGAACT 211451 CACTATAGTC ATTCATATTA GGATATTGCA GTTTCAGCTG ACTTTTCCAA 211501 GTCACTTCGA TTAAGTTAAT GCAAATACCA TAGGCAATAA CCAAGAGAGC 211551 TAATAAAAGA ATATAAGGAG ATCTAGCAAG ATAGAGGAAG CTATCTTTCA 211601 TATTCATTTT AGGTTTAGCA CCTTTTTTCC CCTTTTGCAT TTCTTCTGGA 211651 TTATAGAAGC GAGGATCGGT CAATACGTTC TTATTGATCC ACCAGTAACT 211701 GGCCATAAGA ACAAGTCCAG ATACAATAGT CATAGCCATC AAAAGACGTA 211751 AAGAAATTCC CCAAGGATCT ACACCTTCAG AAACGGAAGC TCTCAACTTT 211801 GAAGCCCAAA CAATTGCACG ACCAGAAGCT AGTAAAGAAA TATTAGCTCC 211851 GATACCGAAA AGAGCGTAGA AACGCTTTGC TTCGTGGATT TTTGTAATTT

211901 CATTAGCAAA TCCCCAGAAC ATTAGAGATA GCATGACGCT TCCCCATAGT 211951 TCAGCAAGTA CATAAAATGC AGCAAATGTC CAGTTTCTTA AGATGGCAAC 212001 GAGTCCTAGC AATCCTGGAG GTAGGATGGC CTGTAAACGG TCAGCAAATT 212051 CTGTAGGATG TAAAACATCG CGTAGCGGAT AAATTACAGT CGGGAACAGG 212101 GCAAAGAAA TTAAAAAGGG CGTTCCCACT GCATAAAATA AGGCCTGCTT 212151 ACTTAAAATA TTACTTAGCT TTGCATAAAT AAGCATAAAG ATAATAGCAC 212201 AGGGGACAAC AAGCCAAAAC TTGATGAAAG GTATTGCCTC TGCACCAGAA 212251 CCAGGAGCTC CCACAATAAG AGTGTCTTTT GTATCGCGTA ACACCGTATA 212301 GTTAAATGTA ATACAGAAGA ACATTAGGAA CATTGGCAGA ACTTTCTTTA 212351 GCTCGTGAGT ATGTATCGGC CACAAGAAG AGCGCAATTT TCCAAAAGGT 212401 TTTTCTTCGG TTTTTGTCAT ATTTACCCTC TGAAATACTT TTATTTTCTA 212451 TTCTGATAGT TTTTTATTTC GTTGTGGACT AACTTGAATT GTTTCCATTT 212501 CGAAAGACCT TATAATATAC CATCTTTTCT CTCTATTGAC AAGAGGAAAG 212551 GGATACTCCT AAGGGATAAG AGTTCTTAGG CGACTTTATT GCTAAGTCGC 212601 CTGTAAAATT TATTTTTTAA TTAAACGAAA GGAGAGGTAA CTTCAAGGTT 212651 GGGAGAGCTT ACAGCCTTCA CAGCTTCGAC GTAGGCATTC CATAGCTCTA 212701 CGCTCACGAT ACCGTTGCGA ATCGAACGAA TTAACTCCCT TTTTAAAGAA 212751 GTTAACGGTT TTTTAGGGCT AAAGCGAGCG AGGAGGTCTT TGTCTCCAAC 212801 TTGTCTTGCC AGTTCTAAAA CGCGTTCAAT ATCTGTTTTA GTAAAGTTAA 212851 TAAAGTCACG ACGTTTTTTA GAGCCTCTAG GAGCTCTAGG CTTAGGATTA 212901 ATTGTGCGTG GGGTTTCCGA TTCTAAAATA AATGTTTTTA ACATTTTTAA 212951 AAGTTGTTCG GGAGCAGCGC TTTTCATCTT TTTTAGAGTG AAATGATGCA 213001 TATAACCGCC GCTAGGGCCT GGAAGATAGC GACATAAATC GTTTTCTTTG 213051 TTTCCGCAGA CTCTTTTAAT AGCTTTTCCG ATCAGTTTTT CTATTTCTTC 213101 TTTTATATGT GATTGTTCTA CGGCCATGAA ACCCACCTTT TATAATTGAT 213151 GCTTACATAT TTTACAAGAA AGAATGTGGC ATACTTTCAA TTTAATGTTT 213201 AGAGAAATAA AATTAAAGCG CTGTTTTAAT CAATCAGCGA ACATAAATTA

213251 TATTGCAAGT TCTCCTTTTT GCAATATTTT TCCTCTCAGG ATCTTTTTGT 213351 CTATAACATT TTGTTTTCTT ATAAAACAAA CGGGTTTTCT AATTTTTTTA 213401 AAACTATCCT GAACAAAGAT ATTTCTTTCA TTTCTATTAG TTGTATTTCT 213451 TGCTGGATAA AATATATTTT GAATAGAACT TGTTTTTCTG GTACTTTAAG 213501 TGGTTTTTAA TGGATCTTAA AAAATGCTTC TAGAGAGATG GATGCGAAAA 213551 TGGGATATAT ATTTAAAGTG ATGCGTTGGA TTTTCTGTTT CGTGGCATGT 213601 GGTATAACTT TTGGATGTAC CAATTCTGGG TTTCAGAATG CAAATTCACG 213651 TCCTTGTATA CTATCCATGA ATCGCATGAT TCATGATTGT GTTGAAAGAG 213701 TCGTGGGGAA TAGGCTTGCT ACCGCTGTTT TGATCAAAGG ATCCTTAGAC 213751 CCTCATGCGT ATGAGATGGT TAAAGGGGAT AAGGACAAGA TTGCTGGAAG 213801 TGCCGTAATT TTTTGTAACG GCCTGGGTCT TGAGCATACA TTAAGTTTGC 213851 GGAAGCATTT AGAAAATAAT CCCAATAGTG TCAAGTTAGG GGAGCGGTTG 213901 ATAGCGCGTG GGGCCTTTGT TCCTCTAGAA GAAGACGGTA TTTGCGATCC 213951 TCATATCTGG ATGGATCTTT CTATTTGGAA GGAAGCTGTC ATAGAAATTA 214001 CAGAAGTTCT CATTGAAAAG TTCCCTGAAT GGTCTGCTGA ATTTAAAGCA 214051 AATAGTGAGG AACTTGTTTG TGAAATGTCT ATTTTAGATT CTTGGGCGAA 214101 ACAATGCTTG AGCACAATTC CTGAAAATTT ACGGTATCTT GTCTCAGGTC 214151 ATAATGCGTT CAGTTACTTT ACACGTCGCT ATTTAGCTAC TCCTGAAGAA 214201 GTGGCTTCCG GAGCATGGAG GTCTCGTTGT ATTTCTCCTG AGGGTCTATC 214251 TCCAGAAGCT CAAATCAGTG TTCGTGATAT TATGGCGGTT GTAGATTATA 214301 TTAATGAGCA TGATGTCAGT GTGGTTTTCC CTGAGGATAC TCTGAACCAA 214351 GATGCGTTGA AAAAAATTGT TTCTTCTCTG AAGAAAAGTC ATTTAGTTCG 214401 TCTAGCTCAA AAACCATTGT ATAGTGATAA TGTGGACGAC AATTATTTTA 214451 GCACCTTTAA ACATAATGTC TGCCTTATCA CAGAAGAATT AGGAGGGGTG 214501 GCTCTTGAAT GTCAAAGATG AGACTTTTTG GTCTGTACAC AACCTTTGTG 214551 TAAACTATGA GCATGCAGCC GTTCTTTATC ACATATCCTT TTCCTTGGGA

214601 AAGGGGTCAT TAACTGCTAT TTTAGGTCCT AATGGAGCTG GTAAAAGCAC 214651 TCTCTTAAAG GCTTCCTTAG GCCTGATCAA ACCCTCTTCG GGGACTGTTT 214701 ATTTTTTAA TCAAAAATTT AAGAAGGTGC GTCAGCGCAT AGCCTATATG 214751 CCTCAGAGAG CTAGCGTGGA TTGGGATTTT CCAATGACTG TCTTAGATTT 214801 AGCCCTTATG GGGTGTTACA GCTATAAAGG AATGTGGGGG AGAATTTCTT 214851 CGGATGATCG AAGGGAGGCC TTTCATATTT TAGAAAGAGT TGGTTTGGAA 214901 TCCGTAGCAG ATAGACAAAT AGGACAGCTC TCAGGAGGAC AGCAACAAAG 214951 AGCATTTTTA GCACGTGCTT TGATGCAAAA AGCAGATCTA TATCTTATGG 215001 ATGAGTTGTT TTCAGCGATT GATATGGCTT CGTTTAAAAC ATCTGTAGGG 215051 GTTTTGCAAG AGCTGCGAGA TCAGGGAAAG ACTATCGTCG TTGTTCATCA 215101 TGACTTGAGT CATGTGCGTC AACTATTTGA TCATGTGGTT TTATTGAATA 215151 AGCGTTTGAT TTGTTGTGGC CCTACTGATG AATGTCTGAA TGGAGACACT 215201 ATTTTCCAAA CGTATGGTTG TGAAATTGAA CTTTTGGAAC AAACCCTGAA 215251 GCTCTCTGA GGAAAACAAT TTGGATCGTG CTGATTATGC TCAGTTGTGT 215301 TTTTTCTGAT ACGATTTTCT TATCTAGTTT TTTAGCTGTC ACTTTGATTT 215351 GTATGACCAC AGCTTTGTGG GGGACAATTC TCTTGATTAG CAAGCAGCCT 215401 CTTTTAAGCG AAAGTTTATC TCACGCGTCG TATCCAGGAC TTCTAGTTGG 215451 AGCTTTGATG GCGCAATATG TTTTCTCATT GCAAGCTTCT ATTTTTTGGA 215501 TTGTGTTGTT TGGGTGTGCT GCTTCGGTAT TTGGTTATGG GATCATTGTT 215601 TCTTGTGGTA TTCTTTGCTA TCGGAGTGAT TTTAGCCAGT TATGTCAAGG 215651 AAAGTAGCCC TACGCTATAC AATCGCATTA ACGCCTATCT ATATGGGCAA 215701 GCAGCCACTT TAGGTTTTCT TGAAGCTACG TTGGCTGCGA TCGTCTTTTG 215751 TGCTTCGTTA TTTGCTTTAT GGTGGTGGTA TCGACAAATT GTTGTGACTA 215801 CTTTTGATAA AGATTTTGCT GTTACTTGTG GCTTAAAGAC TGTTCTTTAT 215851 GAAGCACTCA GTCTAATTTT TATATCGTTG GTGATCGTAA GTGGAGTTCG 215901 AAGCGTAGGG ATTGTTTTAA TTTCTGCTAT GTTTGTGGCT CCTTCTTTAG

215951 GTGCTCGTCA GCTTTCCGAT CGTCTAAGTA CAATTCTTAT CCTTTCTGCA 216001 TTCTTTGGAG GGATTAGCGG AGCTTTAGGA AGCTATATCT CTGTAGCATT 216051 CACATGTCGT GCTATTATAG GGCAACAGGC GGTGCCTGTA ACCTTGCCTA 216101 CGGGACCTTT GGTTGTCATT TGTGCTGGAT TATTGGCCGG TCTATGTTTG 216151 CTTTTTCTC CAAAATCTGG GTGGGTCATT CGTTTTGTCC GTAGGAAGCA 216201 CTTTTCGTTT TCAAAGGATC AAGAACACCT TTTAAAGGTG TTTTGGCATA 216251 TTTCTCATAA TCGTTTAGAG AACATTAGTG TTCGAGATTT TGTCTGTAGT 216301 TATAAGTATC AGGAGTATTT TGGGCCTAAG CCTTTCCCTA GATGGAGAGT 216351 TCAGATTTTA GAATGGCGGG GTTATGTTAA AAAAGAACAA GATTATTATC 216401 GACTCACAAA AAAAGGAAGA AGTGAGGCCT TAAGATTAGT TCGTGCTCAC 216451 AGATTATGGG AATCGTATCT TGTGAATTCT TTAGATTTTA GCAAGGAAAG 216501 TGTTCATGAG TTGGCTGAGG AAATAGAGCA TGTTCTTACT GAAGAATTGG 216551 ATCATACCTT GACAGAGATT CTCAATGATC CTTGTTATGA TCCTCATCGA 216601 CAAATTATCC CAAATAAAAA AAAGGAAGTC TAATGGCTTT GGGACCTTCT 216651 CCTTATTATG GAGTATCTTT TTTCCAATTT TTTTCAGTAT TTTTTTCGAG 216701 ACTGTTTTCT GGAAGTCTTT TCACGGGTTC TCTCTATATT GATGATATTC 216751 AGATTATAGT ATTCCTTGCT ATTTCCTGTT CAGGTGCTTT TGCAGGAACT 216801 TTTTTAGTCT TGCGAAAGAT GGCTATGTAT GCGAATGCTG TCTCTCATAC 216851 TGTCCTTTTT GGTTTGGTCT GTGTTTGTTT GTTTACGCAT CAACTGACGA 216901 CCCTCTCTT GGGTACCTTG ACTCTTGCAG CAATGGCAAC AGCTATGCTG 216951 ACAGGGTTTC TTATTTACTT TATTCGTAAT ACTTTTAAAG TTTCAGAAGA 217001 GAGCAGCACC GCTCTAGTCT TTTCTTTATT ATTCTCTCTG AGCCTTGTTT 217051 TGTTAGTCTT TATGACAAAG AATGCTCATA TAGGAACGGA GCTTGTGTTA 217101 GGAAACGCAG ATTCTTTAAC GAAAGAGGAT ATTTTCCCTG TCACTATTGT 217151 GATTTTGGCT AATGCTGTAA TTACTATTTT TGCGTTCCGT AGCTTAGTTT 217201 GTTCTTCTTT CGATTCTGTA TTTGCCTCTT CTTTAGGAAT TCCTATTCGG 217251 TTGGTTGATT ATTTGATTAT TTTTCAACTT TCTGCATGTC TTGTAGGAGC

217301 TTTTAAGGCT GTAGGTGTAT TAATGGCACT TGCTTTTCTG ATCATTCCAT 217351 CGCTTATTGC TAAGGTTATT GCAAAATCGA TAAGGAGTCT TATGGCTTGG 217401 TCGTTAGTTT TTAGTATTGG TACAGCATTT TTAGCTCCTG CATCTTCGAG 217451 AGCAATTCTT AGTGCTTATG ATTTGGGGTT ATCGACTTCG GGAATCTCTG 217501 TAGTGTTCTT GACGATGATG TACATCGTGG TTAAATTTAT AAGCTATTTT 217551 CGAGGCTATT TTTCTAAAAA TTTTGAAAAA ATAAGTGAGA AAAGTTCTCA 217601 ATATTAGCAG TGATTTAAAA GAATGAAATT TTAGAGTTTA GTCCACTTTA 217651 GTTAGAAATA GTATCAATAG AGAATTGACA ATTCCTCGAC TTGCGGAGTA 217701 TGATTCTCTT TTTCACCTAG TTAAAGGTAG CATGCTTGAA ACATTTAGCC 217751 GTTCTTGGGT CAACAGGTAG TATTGGCCGT CAAACATTAG AGATTGTGCG 217801 GCGCTATCCT TCAGAATTTA AAATTATTTC TATGGCTTCT TATGGAAATA 217851 ATCTAAGGTT ATTTTTCAG CAACTAGAGG AGTTTGCTCC GTTAGCCGCA 217901 GCGGTCTATA ACGAGGGT TTATAACGAG GCCTGTCAGC GATTCCCCCA 217951 TATGCAATTT TTCCTAGGCC AGGAGGGTTT AACCCAACTT TGTATCATGG 218001 ATACAGTCAC TACTGTCGTT GCTGCTTCTT CAGGAATCGA GGCGCTACCC 218051 GCGATTCTAG AGTCGATGAA AAAAGGAAAA GCACTAGCTT TAGCAAACAA 218101 AGAAATTTTA GTTTGTGCTG GCGAATTGGT TTCTAAGACT GCAAAGGAAA 218151 ATGGTATAAA AGTTCTTCCT ATTGATAGCG AGCATAATGC TTTGTATCAA 218201 TGTTTAGAAG GCAGGACGAT TGAGGGAATC AAGAAACTGA TTCTTACAGC 218251 TTCTGGAGGG CCTCTGCTCA ACAAGTCTTT AGAAGAGCTT TCTTGTGTAA 218301 CAAAACAAGA TGTTTTGAAC CATCCTATAT GGAATATGGG TTCAAAAGTG 218351 ACTGTGGACT CATCCACATT GGTCAATAAG GGACTCGAAA TTATCGAGGC 218401 GTATTGGCTG TTTGGTTTAG AAAATGTTGA AATCCTGGCT GTAATTCATC 218451 CTCAGAGCTT AATCCATGGT ATGGTAGAGT TTTTAGATGG GAGTGTGATT 218501 TCTATCATGA ATCCGCCTGA TATGCTCTTC CCAATACAAT ACGCTTTAAC 218551 AGCTCCAGAG CGTTTTGCAT CTCCTAGGGA TGGTATGGAT TTTTCGAAGA 218601 AACAAACTTT AGAATTTTTT CCGGTAGATG AGGAGCGATT TCCTAGTATC

218651 CGTTTAGCAC AACAGGTATT AGAGAAACAG GGGTCTTCTG GAAGCTTTTT 218701 TAATGCAGCC AATGAAGTAT TAGTGCGGAG GTTCCTTTGC GAAGAGATTT 218751 CTTGGTGTGA CATTTTACGC AAATTAACGA CTCTTATGGA ATGTCATAAG 218801 GTTTATGCCT GCCACTCTTT AGAAGATATT TTAGAAGTAG ATGGTGAGGC 218851 TAGAGCTCTT GCTCAAGAAA TATAATCGAG TAGGTATATG ACAATAATCT 218901 ATTTTATTCT AGCAGCCCTA GCTTTAGGGA TTTTAGTGTT AATTCATGAA 218951 CTTGGTCATC TGGTAGTAGC AAAAGCTGTA GGAATGGCTG TAGAGAGTTT 219001 TAGCATAGGC TTTGGTCCTG CTTTATTTAA AAAGCGTATA GGCGGCATAG 219051 AATATCGCAT TGGATGCATT CCTTTTGGAG GCTATGTTCG TATCAGAGGT 219101 ATGGAACGTA CCAAAGAAAA AGGGGAGAAG GGGAAGATAG ACTCTGTCTA 219151 TGATATTCCT CAGGGATTTT TTAGTAAGTC TCCTTGGAAA CGCATTCTGG 219201 TTCTTGTTGC TGGTCCTCTT GCCAATATTT TATTAGCTGT CTTGGCTTTC 219251 AGCATTCTTT ACATGAATGG GGGAAGAAGT AAAAATTATA GCGACTGTTC 219301 TAAAGTGGTA GGTTGGGTCC ATCCTGTTTT ACAGGCAGAA GGATTGCTCC 219351 CTGGAGACGA GATTCTTACG TGTAATGGTA AGCCTTATGT GGGAGATAAG 219401 GACATGCTAA CAACCTCTTT ATTAGAGGGG CATCTCAATC TAGAAATCAA 219451 ACGTCCTGGC TATTTGACAG TTCCTAGCAA AGAGTTCGCT ATTGATGTTG 219501 AGTTTGATCC CACAAAATTC GGGGTTCCCT GTTCTGGAGC GAGTTATCTT 219551 TTGTATAGCA ACCAGGTGCC CCTAACGAAG AACTCTCCTA TGGAGAATTC 219601 AGAGCTACGT CCGAATGATC GTTTCGTTTG GATGGATGGC ACACTTCTTT 219651 TCTCAATGGC TCAGATATCT CAGATACTCA ATGAGTCTTA TGCTTTTGTG 219701 AAAGTAGCAC GGAATGACAA AATCTTCTTT TCTCGTCAAC CTAGGGTATT 219751 GGCTTCCGTT TTACATTACA CTCCCTACCT TCGTAATGAG CTTATAGATA 219801 CGCAGTATGA GGCTGGACTT AAAGGCAAGT GGTCTTCGTT ATATACATTG 219851 CCTTATGTAA TCAATAGTTA TGGATACATA GAAGGTGAAC TTACTGCTAT 219901 AGATCCAGAG TCTCCTTTGC CACAACCTCA AGAGAGGCTA CAGCTTGGGG 219951 ATCGCATTCT AGCTATTGAT GGAACTCCTG TTTCTGGAAG TGTAGATATT

220001 TTACGTCTTG TTCAGAACCA TCGGGTCTCT ATTATTGTTC AGCAGATGAG 220051 TCCGCAAGAA CTTGAAGAGG TGAATTCTCG AGATGCTGAT AAGCGGTTTA 220101 TCGCCTCTTA TCATTCCGAA GATCTGTTAC AAATTTTGAA CCATTTAGGA 220151 GAGTCTCACC CAGTAGAAGT CGCGGGTCCT TATCGTCTTC TTGACCCTGT 220201 TCAGCCTCGT CCTTGGATTG ATGTTTATTC TTCGGAGAGT TTGGATAAAC 220251 AGTTGGAAGT AGCTAAGAAG ATTAAGAACA AGGATAAACA AAGATACTAT 220301 TTGGAGCGTC TTGATGCTGA GAAGCAAAAA CCATCTTTAG GGATTTCTTT 220351 GAAAGATCTT AAGGTGAGGT ATAATCCTTC ACCTGTGGTT ATGTTATCAA 220401 ATATTACTAA GGAAAGTTTG ATCACCTTGA AAGCTTTAGT TACTGGACAT 220451 CTGAGTCCAC AATGGCTTTC AGGACCTGTG GGTATTGTGC AGGTTTTACA 220501 TACAGGATGG TCGGTAGGGT TTTCTGAAGT GCTCTTTTGG ATCGGTCTAA 220551 TTAGTATGAA TTTGGCTGTC TTGAATTTGC TTCCTATTCC TGTTTTGGAT 220601 GGAGGTTATA TCCTTCTATG TTTGTGGGAG ATAGTCAAAA GAAGACGTTT 220651 GAATATGAAG ATTGTGGAAA GGATTTTGGT TCCGTTCACT TTTTTATTGA 220701 TCATCTTCTT TATTTTCTA ACTTTTCAGG ATTTATTTCG TTTTTTTGGT 220751 TAAGGCTCCA TTTTTTTGGA AATCCGTAAA GGTTGTAAGG ACGAGAACCA 220801 AAGCAGTGTA TAACGTAGAT GTCTTGGTTA GGGTAAGGCT TAAGTCTTGA 220851 GTTCTTTGT TTGTCTTCCA TAACAGTTTG TTCTAAAGCT GCTTCGGGAA 220901 CGTGCTCGAT TATCGGGGTT GTATTTCCTT CTTTTAAAAA CTCTTTCATA 220951 AGATAGAGCG ACTTCCGATT AGTAATGATG TAGGTATGAC TCGTGTGGAT 221001 CACGTAAGCT TTATGTGGAG TATCTAGCTT TGTCTCTATA GTGCATACCT 221051 TACCATCGTT TTCATAGGGC AGGCGGAAAG GAAGGAATTT GCTATGTCTG 221101 TTCCCACTAA GAATGAGAAC GTCTAAAGAC GAAGGGAGTT TCCCGACATT 221151 TAACATCTTT GTGGGGCAGT AGGTAAGAAG CTGTCGTCCT AATTTTCCTC 221201 CAAATACGAA CTGTACGAAT TTCACACAGC GGTAGCGTCT AGCTAGTGTA 221251 GACCCTGCGT TCGGAGGAGC CATGAGAATA GCTTTTCTTT TTTGGCTTCT 221301 TCGGGGCAAT CAGGGTGAGC GAGTGCTACA CGAACAATGA CTCCTCCAAT

221351 AGAATGAGTT ACGAAGTTTA TAGGGACTCC AGGCTTAAGT TCAGCTATTT 221401 TTTTCAGCAA GCGATTGAGA TGTTCAGCAT GCTTTTCTAG AGTAAACTTG 221451 CGCGTCTCAT AATTCCAAAT AAAGACATCG TAATGTTCTT TTTCTAGAAC 221501 GCGAGCAATA GGTTTTAAAG ATGTATAAGA TCTTAAAAAC GCATGCACGC 221551 AGACCACGGA TTCTTTTGT TTTGAGGTTT CCTTTAATCC CCCAATTCCA 221601 GATGGAAGGG TTTGGATTAC CGAGGTTTCC GAGAATAAAG GATTACCCAC 221651 AGCTAAAAAG AGTATAGTTA ATAAAAATTT CTTCATGTGT CACCGATATA 221701 ATTATTTAA ATAAGTTTTT TATTAAAAAA GAATATGTTT ATTAAAACTT 221751 ATTTATAAAA ATAACTAATT CGCTTACTTT TAACGCAAAA GAAAAATGAT 221801 GTTATTGTTT TAGATAATTT GCTCAGAAAC TTGAGCGTTC TCGATACTTA 221851 AAACAAGGCT TGTTTTTGGA AGTTCCCCAT GCATATGGGT GGAAGTAATC 221901 AGAGTCTGAC CCAGAGTTGG GGCAGGGTCA AGGAGTTGAC CGACACGTTC 221951 ATTATCTAAT CCAGCATGGA TATCATCTAG ACAGACTAGA GGGGAGACGT 222001 GATGAGATTG CTTTAGATAT AGGCACTCAG CAAGCCTTAA GATTGCCAAA 222051 AGACTGTGTT TCTGCCCTTC ACTAGAGAAT TGAGACACAG GCATTTGGTT 222101 CATAGTGAGT AGAAAGTCTT CGCGATGAGG GCCAACGGAA GTGCTTCCTC 222151 ATTCGAGATC TCTAGGAAGT GATATAGAGA GCTGTTTATG AAATTCTTCG 222201 GCAACAGCAG TTTCAGAAAT ATCAGAATTT TTAATTAAGG AACTTTTAAA 222251 TTTTAAGGCC AATTGTTCTT TTAGGTTGTT AGACCAAAGT TCTTTGGATA 222301 AATCTGAAAG TTTCTGACTA CAGAGAAACC GTTGGATGGA TAGGTAGGTG 222351 CCGTGTTTGA CCAACTGTTC ATCCCAGAGG CCACGGTTGA GGTTTGCTTG 222401 CTTTTTAAGA GAGCATTTCT CTGCTGAAGA GCGCGATGAT AGTACGATAA 222451 GCAGAGGGTA TAGTGGTTAT CGCATTGAGA TAAAAGCAGA TTTAGGAAAA 222501 GGCGACGATC CGCAGGAGCT CCTGAAATTA GAAGGCGGTC TTTTGAAGAG 222551 AAAAGCACAA TAGGTACTTT CCCTATCAGC TGCGATAAGG TTTTTATAGG 222601 AAGTTGGTTA TAGCAGATTT TTTTTCCTTG CTTGTCTGTA TAGATGGAGA 222651 GAGCTTGGGG AAGGTGGTCT TTCTCAAACT GTGTTTCTAA GAAGAAATGG

222701 GAAGATCCGA AGGTGATGGT ATCTGTGAGA TGTTGCGTGC GAAAAGACCT 222751 TCCCAAGGAC AAAACATAAA GCGCTTCTAG GAGGTTTGTT TTTCCTTGGG 222801 CATAATTGAG TTTAGGAGCC AGTGAGATTT CTAAATCACT GTGGTTACGA 222851 AAATTTTTA GCTTCAGAGA GCAGATTTTC ATAAACATCG GCAGGGTAGT 222901 GACCTAAAGG ATTCTCCTTA GGGAGTTTAT TAATCATCAT GTAGCCTCAT 222951 AGGCATGATG ACAAATAATC CTGAGGCAGA ATCGGTAATG ATTCCAGGAT 223001 TATAGGAATC CGAGATCCCT AAGCTGACTA ATTCATCCTT ACTATGCTTC 223051 AGGATATCTA AAAAGAAAAA GGGATTAAAG GCAATTTCTA GGAGTTCGCC 223101 AGAATAATTT ACAGCCATGC TTACCTTTCC TTCACCCACC TTAGTACAGT 223151 TGGCTGTTAG AGTGAGCTCT CCGGGTAAGA AAGAAAACTT CACGGAGTGA 223201 GAGGACTCAT TTGTAAATAA AGCCACTTGT TTGAGCAGAG TAATTAGTTC 223251 TTCGCGATGC AGATCGAGTT TTACGTTGCT TTCTGTAGAT ATGACGGGGG 223301 AGAAATCTGG AAATTCTCCA GAAAGAAGTT TTGTGATCAG GAGAGTATTG 223351 TCACATTCAA CCGCAATCTT ATCTTGATCC AAGAAGATCG CAGCTTCACC 223401 TTCATCGGAG CACATCTTTA TAATTTCTTC TACTGCTTTG ATAGGAATAA 223451 TATATTCCCC AGAAAAACTT TTATCTAAAG TAACTTCAGC ATCTATTTTT 223501 GCTAAACGCT TTCCGTCAGT CCCTACGATG GTAGCCACGC CATTGGCGAT 223551 AGCAAGCAGG ACTCCAGTAA GAACATAGCG GCTTTCTTCT CTAGATACAG 223601 CGAATGAAGT TCTCTGTAGC ATGGTTTTTA GCTGCTCTGC AGGCAAGGAA 223651 AAACGCAAAG CATTTTGTAT ATCAGGGAGC ATGGGGAAGT CTTCTTTTTC 223701 CATGCTGAGT AGGCGAAAGC ATGAAGATCC CGAGGTGATT TGTGCCATTT 223751 CCCCTGCTGA AGAGGAAATT TCTAAATTTG CCTCTGTTAA TTCTTTTACT 223801 AATTGAAAAA ATCTCTTGGA GGGAATGGAA ATAGCGCCTT TCTCATAGAC 223851 TTTAGCTTTG GTGACGCAAC GTGTGCTCAC TGTCAGATCC GTAGCAGTGA 223901 AAACTAATTC ATCATTATAA GTTTCAATCA AAACATGGGT GAGTACTGGA 223951 ATAGGTGTT TTTGAGGGAC GACACTTTGA ATTTTTTTGA TAAGGTTTCC 224001 TAGCTCATTT CGGGATACAA CGAATTTCAT ATTTTCCTAT AACCTGAGTC

224051 TAACTTTATA TGAAGGGCTA CGCCTCTCTA GGAGAGACGG TAGAGGTACG 224101 ATCGGGATAT TACTACAAAG TGATTTTCTG GTCTTCTAGA TATTTTACTA 224151 ATAAAAAGG CTTTAGATTG AGGAAATCTT CCCTGGAAAT CAAGGAAAGA 224201 GGATTCTAAT CATTGTCTTA AGAAGGAAAA ATTGCTTTTT GTTATACTGG 224251 TTTTTGTCCC CAATTATGGG GGAGGATCTT ATGGCACAAA AAGAAATTGT 224301 TTCTAATCGC AAGGCTCTGC GTAACTATGA AGTTATAGAG ACTTTAGAAG 224351 CAGGCATCGT TTTGACTGGG ACTGAGATTA AGTCGTTGCG CGATCATGGG 224401 GGAAACCTCG GTGATGCTTA TGTCATTGTT TCTAAAGGTG AGGGGTGGTT 224451 ATTAAACGCG AGTATTGCTC CCTATCGGTT TGGAAATATC TATAACCATG 224501 AGGAGCGTCG TAAACGTAAA CTCCTTCTTC ATAGATATGA ACTTCGTAAG 224551 TTAGAGGGTA AGATTGCTCA AAAGGGCATG ACTTTGATTC CTCTGGGAAT 224601 GTTTCTGAGT CGCGGCTATG TTAAGGTACG TTTGGGTTGT TGTCGTGGGA 224651 AAAAAGCTTA TGATAAGCGT CGTACGATCA TAGAAAGAGA AAAGGAACGT 224701 GAAGTTGCCG CTGCTATGAA GAGGCGCCAT CATTGATATA GGTTAGGATA 224751 TGGTGTTCTT CAGCCCACTG TTTTGCTTCT ATTTTAGAAT CAAAAGTCAT 224801 GAGGACTGTG GCAATAGCGT CGGCGTATGC GCAGCTCGGA TGGACTACTG 224851 AAACACTTTG GATAGGATAG GAGCTTAGCT CTAGGGGTTT CCCTGTACGA 224901 GTATCAAGAA TATGGGTGTA AATTTTTCCT TCAACACCC ATTTTTGAAT 224951 ATGATTTCCA CTTGTTGCAA TTGCCATATC ATCGATATCT AAGATCGTAC 225001 CTGCTGCTTC AGAAAAAATA CGCCAAGGTC TTCCCGAGGG ATGATGCCCT 225051 GACGTTTTGA TCTCTCCTCC CCACTCTACA TAGTTGTTCG GACAAAAGGT 225101 ATTGCAAATT TCATTTAGAC AATCTACGGC ATAACCTTTG ACAACACCAC 225151 AGAGGTCGAT TTGAACATGA GGATTCTTTT TGATTAGAGT TTTTGTGTTT 225201 GACTGAAACT CCAAGTGTTG CCAGCCCATG TCTTTATAAT GTTGTTCCCA 225251 AACGTCTTTA GGGGGGAGGG TTTGACTTTT GAGATGTAGA AGCCATAGGG 225301 TTTTTAAAGG TCCTACAGTA GGGTCAAAAC GTCCTTCTGA AAGTTTGTAA 225351 AGTGTATCTA CCTGATCTAG AAACTCGGAA AGTTCTACAG ATAAAGTTAT

225401 GGGGACATCT GCTGGAGCTC GGTTGATTAT CGAGAGTTCA GAATAGGGAT 225451 TCCAGTTGTT ATAAATCGAG TCGATCTTAT GAAAGCATCT ATCAATTTGT 225501 TGGGATAAAG ATGCTTTTC TTTTGCGGAT AAAGAGGTTC CCAGAACAAT 225551 GCGATAGAAG ATTGTCATCT GCTCTCCTTC GATTGTTGTC GTTTTTTGAG 225601 AGCATGAACA GAGTCCAAGA CATAAAAGAA CTAAGAAAAA TTTTGGTAAC 225651 ATCGCCATGT ATTGCTCATG AGCATAGCGA CAGTCATGGG ACCAACGCCT 225701 CCAGGAACTG GAGTGATTGC TGCGCATTTT GTCACAACGT TATTAAAATC 225751 TACATCCCCA AGAAGAGTAT AGCCTTTCGC ATTGTCTGCA GGGACTCTTG 225801 TTGTTCCTAC ATCTACGATC ACAGCATGTG GGGCTACCAT AGTTTCCTTT 225851 ATAAAAAGCG GTGCTCCTAG AGCAGCAATA ATGATATCAG CTGTCTTTAA 225901 GATTTCTGGG AGGTTTTCCG ACTGGCTATG AAGAACTGTG ACTGTACAGT 225951 TAGTTTGAGG ATGCTTTTGC ATCATGAGGG CCGCTAAGGG TTTCCCCACG 226001 ATGTTGCTTC TCCCTACAAT AGCGGCATGG CGGCCTCGAA GAGGAATTTC 226051 ATAATAGTTC AGGAGTTCAA TAATTCCTGC AGGAGTGCAG GGTAGAAGTC 226101 CATCAAAATT TCCAAGGAGC AACTTTCCCA TGTTCACAGG GTGAAGCCCG 226151 TCCACATCTT TGTCTGGGGA GATCGCTTGG AGAATCACTT CGCTGTCCAA 226201 GTGTTTGGGC AAGGGAAGTT GCACGAGGAT GCCGTGGATG CTAGGATCTT 226251 GATTCAATCG TTCTATGAGC TTAAGGACTG AGGAGAGGGT AGAGTCAGAG 226301 GGTAACTTGT GCGCTTTGGA GATAATTCCG ATTTCTGTAG CTTTTTTGAC 226351 TTTCATGCCA ACGTACACCT CAGATGCGGG GTCATTGCCA ATCAGGACCA 226401 CAGCAAGCCC CGGAGAGGTA GGACTTTGTG AGATTTCCTC TTTGAGTCTC 226451 TGAAGGATTT TTTCAGCTGC AGGAATCCCT CTCAGTAACA TACCAATCTC 226501 CTAAAAAAA ATATCATTTT ATGGATTCCT GAGTTTATTG CGCAAGGAAG 226551 TATATAGATT ACAGAAGTGA AGATGAGATT ATCTAAGAAA GAACATTTTA 226601 GAGGTATCAT TAGCACATCT ATTGACCAAC GGACTTCATG TCCGAACAAG 226651 GCGCACATAG GAATGGTTAG GAGTAAAAAG ACTCCATAGA AAATCATATT 226701 CACAATGACC AAGGCGGCTT TGCTCTCTTT TGAAAAAATG AGGTGTGCCT

226751 TATGGGTGAT CAGAGCAGTT AGAGGGTAGA GCAGTAGAAA TACACCGCGG 226801 CTTCCTAAGG CAAGATCACT TAAGACACCA CAGCCTAGAG CGAGTACCAG 226851 TCCCTTATCT TTTGAGAAAC AATAGAAACA AAGAACAATA TAAGGGGAAA 226901 AAAGGATCAG ATGTGCTTTA GGGAAGAGCG TTGGAAGGGT GAACATGGAT 226951 AAAATACACA AAATAAAGCA TTTGCGCGGT GTTATGTGGA CAGACATGGT 227001 TTACGCTTTG AATTATGATT ATCGATTTCA AAAGGTTTTG CAGACTTGTA 227051 TATAGCAAAG CATAGGATTC TTAGCTTGAT TTTTTATGTT TATCTAAACT 227101 GCTTTTAAAC AATTATTTAT ATAATATTTT ATTTTTAGAT TAAAATTAGT 227151 TTATTCTTAA TAAGTTTTTT AATTAAAATT AGTTTATTAA GTTGATTTTT 227201 TTGTTTTGT ATTTTTATG GGGAAGCCTA AGAAGAGCAG AACGGATAGG 227251 GCTTTGGCTC AGGAGATTCA AAAGAAATCA ACGGAAGTGT TGAAGAAGCC 227301 TGCGCGGATA AAAGCTAAAA ATCGTCGTAA ATTTCTTATT GCTAAGGAAC 227351 AGAAAACTCT TAAACACCGT GCTCAAGAAT ACGATCAGTT AGTTCGCTCT 227401 CTCTTAGATT CTCAGAAGAA GGACACCGAT AAAGTTTTGA TTTTCAATTA 227451 TGAGAATGGG TTTGTTTTTA CTGACAAGGA CCATTTTAGT AAGTACTCTA 227501 TCCGTCTTTA GGCGGGGTTA AATTAAAAAA AAATCGCAGA ACGGGACTAT 227551 CCCGTGATGG AACGACAGGA AGCCTGGGTA ATTAATTTAG AGTTAGGAAC 227601 TGTCTCGGTA TTAAGATCTC CAAAATCTTA ATACCTAGAC AGTCGGATCC 227651 GAGTGATCCC TCTAGCTCGA CAGCGATGCA CTTTGTTGCA TCTACTCGCT 227701 GTTCTCTGTC CTCCCATAAG TTTTTTTACG CAAGGTGTTT CACCTTGCGT 227751 TTTTTTTGT TTTTTAGATT TTTAAAATTT GTTTGTAAGT TCTTTTTTT 227801 AAATTAATTA TTTTTTGTTT TAAAATTATA AGTAGTTATA AGTTTTATTT 227851 TTTAAATTCA GAGAATCATT ATGGCAGTTT CAGGTGGCGG AGGGGTTCAG 227901 CCTTCTTCGG ATCCAGGAAA GTGGAATCCT GCTCTGCAAG GAGAGCAGGC 227951 AGAAGGCCCG TCTCCGCTAA AAGAATCTAT ATTTTCTGAA ACCAAGCAGG 228001 CCTCCTCTGC TGCGAAGCAG GAAAGCTTAG TGCGTTCAGG ATCTACAGGA 228051 ATGTATGCAA CAGAATCTCA GATAAATAAG GCTAAGTATC GTAAAGCTCA

228101 AGATCGATCA TCAACCTCTC CAAAATCCAA ATTGAAAGGT ACATTTTCTA 228151 AAATGCGCGC TAGTGTGCAA GGATTCATGT CAGGATTCGG ATCTCGGGCT 228201 TCGAGAGTGT CAGCAAAGCG TGCTTCCGAT AGTGGTGAGG GAACATCCTT 228251 ATTGCCGACA GAGATGGATG TTGCTCTAAA GAAGGGAAAC CGTATTTCAC 228301 CTGAAATGCA GGGATTTTC TTAGATGCTT CGGGTATGGG AGGGAGTTCC 228351 TCTGATATTT CTCAGCTTTC TTTAGAGGCT TTGAAATCTT CAGCATTTTC 228401 AGGTGCCAGG AGTTTAAGTT TAAGCTCTTC AGAATCTAGT TCCGTGGCTT 228451 CGTTTGGATC TTTCCAAAAG GCCATAGAGC CTATGAGTGA GGAGAAGGTA 228501 AATGCTTGGA CAGTGGCTCG TTTAGGAGGG GAGATGGTCA GCTCTCTTCT 228551 CGATCCCAAT GTTGAGACCT CATCATTAGT GCGCAGGGCA ATGGCAACAG 228601 GCAACGAAGG CATGATAGAT CTTTCTGATT TAGGACAGGA AGAGGTCAGT 228651 ACAGCCATGA CATCTCCCAG AGCAGTAGAA GGAAAAGTAA AGGTATCTTC 228701 TTCTGATTCT CCAGAAGCGA ATCCAACAGG AATTCCAAAT TCTAATACTT 228751 TAGAAAGGC GGAAAAGGAA GCAGAGAAAC AAGAAAGTCG AGAGCAGTTG 228801 AGTGAGGATC AGATGATGCT TGCACGTGCT ATGGCTGGGC TTCTTACAGG 228851 GGCAGCGCCT CAAGAGGTAT TGAGTAATTC TGTTTGGTCT GGTCCTTCTA 228901 CAGTATTTCC TCCTCCCAAG TTTTCAGGAA CTTTACCCAC CCAGAGATCG 228951 GGAGATAAAT CAAAGCATAA ATCTCCAGGA ATAGAGAAGA GTACGAACCA 229001 TACGAACTTT TCTCCTCTTC GGGAAGGTAC TGTGAAGAGT GCTGAGGTTA 229051 AAAGTTTGCC TCATCCAGAA AGTATGTATC GTTTTCCTAA AGATAGCATC 229101 GTTTCCAGGG AGGAACCTGA AGCCGTTGTT AAAGAATCTA CGGCATTCAA 229151 AAATCCAGAG AATAGCAGTC AAAACTTTCT CCCTATTGCT GTGGAGAGTG 229201 TTTTCCCTAA GGAAAGTGGT ACGGGAGGGG CTTTAGGAAG TGATGCTGTG 229251 AGTTCCTCAT ATCATTTCCT TGCGCAACGT GGAGTGTCTT TACTCGCTCC 229301 TCTACCTCGT GCTACTGATG ACTATAAAGA GAAGCTCGAA GCTCATAAAG 229351 GTCCTGGAGG TCCTCCAGAT CCTTTGATTT ATCAGTATCG AAATGTTGCT 229401 GTTGAGCCGC CAATTGTTCT CCGTTCTCCC CAGCCGTTTT CAGGATCTTC

229451 ACGTCTATCG GTTCAAGGAA AGCCTGAAGC TGCTTCAGTT CATGACGATG 229501 GTGGGGGGG AAATAGTGGT GGTTTTAGCG GAGATCAAAG AAGAGGATCT 229551 TCGGGCCAGA AAGCTTCCCG TCAGGAAAAG AAGGGAAAAA AATTATCTAC 229601 GGATATTTAG GGTTTTAAGT CGGTTTGATG TGATGCGTAT AATTCGTTTT 229651 GATCCTTATG GTGCGCTATC TGCACAAAGC ATAGCTAAAG ATTCCCGTCA 229701 AAACTCTCCT TTAGTAGAAA AAATTTCTGA GGAAATTGCT ACGAATGAAG 229751 CGATTCGGCT TGCCTTGCTA GCTATTGGAG ATCGCGAACA AGAGGAGAAG 229801 AAACAGAGGC ATCGTTATAA GCTACTCGGA CAAAAGCAAG CCAAGGTCTT 229851 GCTTTCTCAG TTGCGTCATG TGCATTTAGA TTTTAAAAAA CTATATTGCG 229901 ATAGTAAGAA AAAAGAAGAT CAGGAAAAAA CAAACAGAAG 229951 CGATCTATTA AAGTTACAAA GAAAAAAAG GGCATCTCTT TAGGGGCTGC 230001 CGCTTCTCAG GCAATTGCAG CAGCAGCAGA AGCTTGGGTA ATTGCTAGAA 230051 ATAAAGGAGT CTTAGAAACT GCCTCCACTC TTTTTTATCA AAAGGATGAA 230101 GAGGCCTAGA CATACTTGAA TCACGAGCTA GGCCTTCTTC TGTGACTATT 230151 TTAGATCATC ACGCTGCTTC TTGCTCTTCT ACTGGAAGAC TTTTTTCTAT 230201 AATTTCTTTT TCTTCATCGT CTACTGTAGG ATGTTCTGAA TGCTTAGCTA 230251 GATCTTTCCA AATCATTCGA AGTTTCTGAT TTTGTTGTTT GGTCAGGTTT 230301 TCCAGTATGA TCAAGCTTTC ATCATTTAAG GAGAATCTTC CTTTAGACCA 230351 ATTTATAGAT CCTGCAAGTA GAGTTTTATT ATCTATAACT GCAAACTTAT 230401 GGTGAAGAGT ACAGGGTGCG GTATTTATAG AAACAAAGTC TTTATTGATA 230451 TTTAATTGTC GTAATTGCTT AAAAGTAAGT TTGCTATGAC TTCTATCAAT 230501 GATAATATCT ACATGGATTC CTCGTTGTTT TGCTTGATGT AAGGCTTGAA 230551 TAATCTCCGA GTGGGTCAGA GCAAACATAG CAACTTGGAT GGTTTTCTGA 230601 GCTGTCTGGA TTTTTTCGAG TACAGCTTGT ATTGCAATTT TACGATCTTG 230651 AGGAAGAACA AAATACTTTC CTGTTTGATC CTTTATAGAA AAGTCTCCAG 230701 AGGTATTTGT GATAATGAGA TCACAGAGCT CCGAGCTATG CATTCCTAGA 230751 ATGAGATTAT TATCTAAACG TAGAGAAAGA TTGGTGTAGT TCGCAGATCC

230801 TAGCCAAGCA TCTTTCTTAT CTATGGAAAG AGCTTTTTGA TGCATCAGTT 230851 TACGCCCTGC TGGAGGTTGC TCGACTAAAG TTACATTGCT GGCTTGCTTT 230901 AAGATTTGGG GAATTTTAAA TTTTTGATAG TAGATCGTAA CTTTGTTTTT 230951 TGCTTGAGCT TGTCGAGTTA AACTCTGTTG GATCTTGGGT TCTGAGAGGT 231001 TATAAATACG TAGGAAGATC TCTTCATCAG CGTGTTCTAT AGCATCGCAT 231051 AGAATTTTAC GCATGTCCTC ATTGCATTGA TTTGAGTAGA TGATAGCTTC 231101 TTCAGACTTT AAAAAAGTCT TAAAAGTGTC ACCACGAGGA GCTCTTGCAA 231151 AAATTCCTAC TAAAATCAAC GTGCTAATAA TAACACAGAT TTTTAATTTA 231201 TCTTTTGTC TTTTATTCAT TTTTTACTCT TAAAATAGCT TTTTTGTTTT 231301 ATAAGATTCT TATTAAATAT AAAAAGTTTC TTTATTGGGG AAATATGCGT 231351 GCCATCGTTC GGAAACCTTT TGTTTTGTAG ATGGGTCGAC CTCTACTTCT 231401 TTAGGATAGG AAGGCTTCAT CAGGGCATCG ATAACGAAGG GAAAGTTGTA 231451 ATTAGGACGG TGAGTAGCAA AATGGCTGTG GAGCGCGTGA AGATCATTTG 231501 CTGGGGCACA TCGTGTGAAG GTCCTCCAGA GAAAATCTTT TTCACTTTGA 231551 ATGGTTTCTC TCAGATTATC GGCAAGGATA ATCAGAGGCC ATGATTTAG 231601 ATCTGGATGG TGAAGGAGAG ATTTAATACA TCGGTCCTCG AGGGATGTTT 231651 CCAACACTAG GCAACCACGA CAAAAGGGAG CGATGTCTTG AACTCCATGG 231701 ATTTTCCTC CCTGATATCC ATGGGGAAGG TCTCGGATGG CTTTTCCTAT 231751 TCCCATGAAG ATTCCCTTGG AGCCCTTATT TAAGCTTGGT CCTGTATAGT 231801 CTAACGTATC GTTTGCAGTT TCTGAGAAAA TAATAAGATC TCGGTCTGGC 231851 TGTAGACGCT CTAAAATGGT TTCTAGAACC ACGGAGAACC TGTCGAGAGG 231901 CACCTCTTGG TCTGTGACCA TTAGGAATTT CGTTAGGGAA AGTTGGCCCT 231951 CTCCAAGAAT TCTAAGAGCT GTGGTTAGAG ATTCTCTCCA ATAGCGTTCT 232001 TTAACGACAG CGGCAGTCAG TGCATGAAAC CCTGATTCTC CGTAACTTTT 232051 AAGTCTACGC ACACCAGGCA TAACTAACGG AAATAAAGGG GAGAGGTATT 232101 CTTGGAGTTT GTTCCCTATA TAAAAATCTT CTTGGTAGGG TTTGCCGACT

232151 ACTGTAGCAG GATAGATTGC ATCTTTTCTG TGATAGATTT TATGACAGTG 232201 GAATTCAGGG AAGTCATGTT GGAGACTGTA GTATCCGAAA TGATCGCCAA 232251 AAGGACCTTC AGGACGACGT TTCCCGGCCG GAGATTCTCC GACCAGGATG 232301 AATTCCGCAT CGTAGAGTAG AGGGTGGGGA TGGTCGTTTG TTTTTTTATA 232351 AAGGAGCTTC GCTCCTTGGA GGAAGGTAGC AAAGAGAAGT TCCGAGACAT 232401 TCTCAGGTAG GGGGGCAATC GCAGAAAGGG TTAAAAAGGG GTTTCCAGAC 232451 AAAAATACCG AAACAGGAAG GTTTTGCTTT TTTTGCTCTG CTTCATACAG 232501 ATGCATCCCT CCGCCCTTCT GGATTTGAAA ATGGAGGCCC ATGGTGTTTT 232551 GATTGAACCG TTGCACGCGA TACATCCCAA GATTGGGTGT AGTAAGAGTC 232601 GGCGATTCCG TATAGACAAG AGGAAGTGTG AGAAAGGCTC CACCATCTTC 232651 AGGCCAGCTT GTGAGTAAGG GAAGGTGATC TAAGTTAACT GAGGACATAG 232701 AAACAAAGG AAAGCGACGG AATCGAGCTT TTTTGAGCCC TAAAGAGCTT 232751 ATTCTTTTTA ATAGATCCCG AGATTTCCAT AGAGAAGAAA GCTTTGGTGT 232801 AGAAGAAATA AGGTGGGCAA CTCGAGCGAT GAGGTTATCA GGAGCTTGAG 232851 AAAAAAGTTG GTCTACACGA TGTTTTGTTC CAAAGAGATT GGTCAGGACT 232901 GGGAATGACG ATCCGATGAC ATTATGAAAA AGAAGGGCAG GGCCTTGATC 232951 TTCAATAACA CGACGATGAA TCTCAGCTAA CTCGAGGTTA GGACTTACGG 233001 GAGCAAAAAC ATCAATAAGT TGTTTTTGTG AACGAAAAAG AGAAATATGA 233051 CGCCTTAAGA AAGACATAAA TATTTCCCTA TACTTAAGTT AAATTAAAAA 233101 TTTTTACTTT TAGCTCTTTC GAGAACTTTC TCTAATCCGA GCTTATCAAT 233151 GTGACGAAGA GCGCTAGCAG AAATTTTAAG CTTAAGAAAA CGGTTTTCTT 233201 CTGTAGACCA TAGACGCTTG GTCAACATAT TAGGGAAAAA TCTTCTTTTA 233251 GTCTTCCCTG TCACTTTCAA ACCAATTCCT TTTTTCTTTT TAGCAATACC 233301 TCGAAGTGTA TAGCTATAAC CACGGCGAGG TCTCTTTCCT GTAAGTGGGC 233351 ACTITCTTGA CATATTCTTC CTATGAATTC TCTAACATCC GCTCAATAAA 233401 GCAACTCTAT GAAAAGGAAT CTATACTAGA GCTTTTCGAA TATATGGGGA 233451 AGAGGTTTTC TTAGAAAATG TGATTGTCTG TTGTTGATAA TGAAGGAATT

233501 CTTTTAAGAA AGACGGTCGT TCCAGGAAAT TATTATAAGG TTCTTACTAT 233551 AAAAAGACTT AAATGTTTTA TTGCTATCCT TACAGTCCTG TAAGGATCTT 233601 CTCAATGTAA CCATTAAATT TTTTATGAAT AGCGAGTTCT TCTAAGGAAG 233651 GCCGAACTCG ATACGACCAA TTCTTTTTAG AAATTGTCCC AGGTGTATTA 233701 ATGCGTTCTC TTTGTAGATT TTTTGATACT AAATCAGGGC AGAGGGCGAG 233751 ATAATCGTTA AAGAGGTTGA TATGAAAGAT AGATGCTGAT TCATGAGAAA 233801 GTTTTAAGAT GTCTATTTGA GTTTCTGTAG TCAGGGTTTT TTGAAAAGGA 233851 AGATGTAGAA ATTTAGCAAA TTGCTTAGCT TCCTTAGGTG AATTGAGCCA 233901 CCATTGGGCA AACGTATCAG AGTCGTGGGT AGAGAGAGTG GTCACAGAAA 233951 GTGGATTATA ATCTTTTAGG GGAATGAAGG CACTGTCGCT TTCCCAGTTG 234001 CGTTCCCATC GTGGAATCCG GGTTCCACAG ATTCCTAAGT GTGTTAATGT 234051 CGTTTTGACG TCTTGGGGTA TAATCCCTAA ATCTTCTCCG ATAGGTAACA 234101 TAGAAGAGC TCCGAGCATA GTAGAAAGGA TCTCCGTGCC CTGCTTTATA 234151 TAGTCTTTAG GATTGTCTGG AATGAACCTT CCTCTTCCTG AAGAATCCCA 234201 AATCCACAAA CGGAAAAATC CTATAATATG ATCTAAGCGA TAGACGGAAT 234251 AGAAGTTTTG AGCATATCGC AGACGCTCTT TCCACCAAAT GTAGTCGTCT 234301 TTGGCAAGTT GTGAAAAATT ATAAATAGGC AGATGCCAGT TTTGTCCTTC 234351 AGAATTGTAG AGGTCAGGAG GAGCTCCTAC AGACCTTGAT GAAGAAAAGT 234401 AGTCTCGGAA ATACCAAACA TCACAGCTAT CCTTGCTAAT AAGAATAGGG 234451 AGGTCTCCTT TAAGCAGGAC GTGGTGTTGA TCTGCATAGG CTTTCACTTC 234501 GCAGAGCTGT TGGTAACAGA GAAACTGTAG ATAGGAAAAA AAGAGGACTT 234551 CATCATGGAA TTTTTTAGTT AAGTCCGGAA AATTCTCCTG ATCTGTGAGC 234601 GACTTCGGCC AGTTATTAAT AGGTTCTCCG TGCATATGAT GTTTGATTGC 234651 ACGAAAGGTC CCATAGGGAT AAAGCCAATA GCGCTCGCTT TCTAGAAACT 234701 CAGAAAATT TGAGTTTCCT TCGAGGGAAG ACTTGCAACA TTTTTGGTAG 234751 TACTCTCTTA AGAATGCCCA TTTTTTTTCT TTAACTTGAG TATAGCTGAC 234801 TGATGGAGTC GAGCATAACT CATGCATATC TTGAAGTTTC TTGGCAACTT

234851 CAGGGATGGT ATCGATATTT GGAAGAGAG ATAGGGAAAG GAATAGGGGA 234901 TTCAGGGCTA CGGAAGAGAT GCTGTTATAG GGACTCGTAT CTTCACCAGT 234951 ATCATTTAAA GGGAGAAGCT GAATAACGCT GAAGCCCTGT TTTTGGCACC 235001 AAGAGATCAG AGGAATGAGA TCTAAAAATT CACCGATTCC ACAGCTATTT 235051 TTTGTGTGTA TTGAAAATAG TGGGAGATAA ATCCCGTGTT TAGGAGAGGT 235101 TCCTATAAGT TTCCAAGCAT GTGCTGAGGG TGAGTGTTTT GTGTATTTTA 235151 AAACATTCAC GCGACGTAGT AGATTCCCAA AACATGAAGG TCAGGAAGAT 235201 TCCCTGTTCT TAGGGATTCC ATCCAAATCT TAGCATGCAA AGAAAAGAGC 235251 TTGAGATACT CGAATAACTT TTCTCCATTT AGGTAGTTCA TACTCAGGAT 235301 ATCTGAAAGA TAGAGCTGTT GAGTGACCTC ACCGTAGCCT AGAATTCCCT 235351 TGATGCTGGA TTGGAACGAG CCATTTACAG AGAGAGCAGC TTTGAAAATA 235401 CGCTCGCGAA ATACGTTTTC AGGAAGAGTA CCTAGTAGTG TCGATACTGC 235451 AAGATCTCCG GAATTTCCAT CTTCTTCTAT TTGCACAGGG ACATGGGTAT 235501 CGCTGAAACG GATTAAACAA GAGTTATTTT TATCTGGAGC AAGTGTCGTA 235551 TTTAATAGGG GTGCTAAGGA TTCTAGTAAT TGCTCGTATT GGTTTTGCAT 235601 GGCAATCCTT TTTTAATCAT GACCAAGGAT AGGGTTTAGG GAAGTCTGAT 235651 GCTTTGGGAT AATCTTCATT GTTTATATTT ACAGCATCTA AAGCATTAGC 235701 AATCATAGCT CCTAATTGCT GACGTTTGTC TGCTGAAGAG AAAAGGCGTG 235751 ACGACGTTTG ACGTAAAGCA GAAAAGAATA AGTTCAAGAC ACCGGTCACA 235801 GAATCAACAT CGTCTCCTAT GAGATTGCGG ACTTCTCGTT CTACTTTAGA 235851 TGCTGTTGGG AACTTATCGT TAATGATTTT ATGGTAGGAC TCAGCTACCT 235901 TCACAAAGTT TAGATCAGAA GGAGTTTGGA TTCCCTCAGC TTTTAAGCTA 235951 TCGAGTAAAA TAGGAACGCG ACTTTCAAAG TAATCGTACG AGGTAAGAAC 236001 TGCTTGCAGG TTACGAGTTT CTGTCATGAG AACTTGTAGT TGCGCACTGG 236051 GTACGTAGGG ACCCTGCCTT TTTAATTCTG TTGCCATTCC TTTCATTAGA 236101 AAGGAGCTGA CAATAGCCAT ATCTTGGTAG GTATAGCGGT CTTGAAGCAT 236151 AGAAAGTAGC TGATCACAGG TATGTGTGTC TCCAGTCACT TCTAAGTACA

236201 AAGAGCGAAG CCCTGAAGGA GAAACATTCA GTTGGTCTGC ATATTCTTGA 236251 GAGGCAAATA AGATGTTTTT CGCACCAATA GCAGTTCGTC CGAATTGCTC 236301 CGTATGAGTA TTCCTTGCTT GGATAAGCGC TTCTTTTAAT TTACCTTGGG 236351 AGGGTGGAGT CGTTTGAACC AGGTAGTCCA AAGCTGTGGA TTGCAGAGCT 236401 GGGTCTTTAA TTTTCTCTTG TACAAGAGCA AGAATGTCTT CTGGAGAAGC 236451 ATCGTCTCT ATTGCATCAC GCAGGCCGCG AAGTTCTTGA CCAGAGATTT 236501 CAGAATTCCC AGAAGCATAC TTATCAGCAA GATCTGTGTC AGGCTTCTCT 236551 TCTGTAGATT CAGATTTTTT CTCAGCCTTT CCAGCTTCTC CTTTTTTCCG 236601 AGATTCTAGA GTTTGAAACT TCTCTTCCTT TTTTTTCGTG CGTGTTGCTG 236651 CTGCGGGATT TGTCAGGTCC TGAGATTGTT GAATCATGTT CATCTCAGAA 236701 CCTTCTTGGC TGGCTACAAC TTCTGCTGCA TCTGCTTTTG CAGCTGCAGC 236751 TTCTACAGCT GCAAGGTTGA CACCCTGAGT GCCTCCTAAA CCACCTGTGC 236801 CTCCTGATGC TGCCATATGC CTCCTATGAG CGACAACGTA TCAATTAGAA 236851 AATCTGAATT CTTCCTAAAG GCTGGATGCG GATTTCTGGT AGGATTTCTT 236901 GATAAGAAAT CACAGCAATG TCAGGGAATT CTGTTTCTAT TAATTTTCGT 236951 ACATATCTTC TTACATCAAT TGCTGTCAAT AATACTGGTG GTTGGCCTCC 237001 TGCAGGTGTT GGCGTGATCG TATTCCTCAT AGATTTTAAA ATTAGGTTCA 237051 CAGAATCAGG ATCTAGAGCA AGGTAAGAAC CTGCCGATGT CTGTTTAATT 237101 GCTCCACGAA TCATCTCTTC AATTTCTGGA TCTAAGAGAT AAACAGAAAT 237151 TGCTGATTGT CCTTGAGAGA ACTTGAAGCT GATATAAAGC TTTAAAGAAG 237201 ACCGTACATA TTCTGTAAGC AAAACTGTAT CTTTCTCAGT TTGCGCCCAC 237251 TCGCTCAGAG ATTCTAAGAT TGTACGTAGG TCTTTAATTG AGATTTGCTC 237301 TTGAACCAAT CTCTTAAAGA TTTCCGTAAG CTTTTGCAAT GGAATAAGCC 237351 TTGTGACTTC CTTCACTAAG TCCGGGAATG AACGTTCCAT AAATTCGATC 237401 ATAGAACGTA CCTCTTGAAT TCCCAAAAAC TCTTGAGAGC TTTTATGGAA 237451 AAAGTACGAA AGATGGAGAA TGATCACTTC GAGCGGCGTC CAATATTTAA 237501 TTGCTGCCTT CTCTAGAATA GCTTTTGCAT CTTCACTAAC CCAAGCTGAA

237551 GGAAGACCCG CAGCATTCTT ATAGGTAATG AAAGGTAGAT TATAACGGCT 237601 GAGATTGTCC TCCACCTCAT TGGTTAACAC ATGGTGCGGA GGAATTTTTC 237651 CTCGCACATA AGGGACTTCA TTAAGCAGAA TCATATAATC GTATCCTTCT 237701 AAAGAAGGG AATCTGTGCG AACATGAATG CCAGGGTATC GGATTCCGAT 237751 ATCCTGATAG AGAGCTTGCC GCATTTTAGG AATCATATCA TCAACAAAGC 237801 TTTGTCCTGA TTTTGTCTTG TGTTGGATAA GCTTAGAGAG ATCTTTTCCA 237851 AGTTCTAGAA TTACGGGAAG AGTTAGAGAA TAGTCATCGG GATTATCCCC 237901 AACAGTAGCA GCGCCATCAC CAGCAGCCCC TACGGTTGTT GAAGCTCCTG 237951 AGCCACCACC TTTTTTCCT GCCGCTGATT TCTTAGTCAG TAGGAGAATC 238001 CCTAAGGCAA CGAAAATTAA TGCTAAAATG GAGAAGGACC ATAGAGGGAA 238051 GCCCTTGAAG AAACCAACCC CTAAAGTTGC AGCACCTGCA AGGAGTAGTG 238101 CTCGTGGTTC TTTAACGAGC TGAGTAGAAA TCTCTTTACC CAAGTTCGTA 238151 TTTTTGTCAC TCGATACACG AGTCGTGACA ATACCCGCTG TCAACGCAAT 238201 CAAAAGAGAA GGAATTTGAG AGACTAAACC ATCTCCAATG GAGAGAAGAG 238251 TGTAGACGTG AGCTGCTTGA GCGAGGTCCA TGCCGTGCAT AGCCACCCCA 238301 ATCGTCAAAC CGCCAACAAT GTTAATCAAA GAGATAACGA TACCAGCGAT 238351 AACGTCTCT TTGATGAACT TCATGGCACC GTCCATGGCT CCGTAGAGTT 238401 CACTTTCCTT TTGGATTTGA GCCCTTTTAT CACGAGCTTG TGTGGCATCA 238451 ATCATACCAG CTCGTAAGTC CGCATCAATC GCCATCTGTT TACCTGGCAT 238501 CGCATCCAAT CGGAATCGGG CAGCAACTTC GGCAACACGC TCGGCACCCT 238551 TAGTTACTAC GATAAACTGA ATGATTGTAA TAATGAGGAA GATAATGAAC 238601 CCGACCACAT AGTTCCCTCC AACCACGAAG TCTCCGAAGG CCTGAATGAC 238651 ATGACCCGCA TACGCTTTAA GGAGAATCTG TCGAGAAGAG GAAATATTAA 238701 TCCCCAAGCG GAACATCGTA GTGATGAGGA GCAACGAGGG AAAAACAGAC 238751 AGCTGCAAAG CACTTGGAAT ATAAAGAGCC ACCATCAATA AGAATACAGA 238801 GATCGATAAG TTGATGGTGA TCATCAAGTC AACGATAGGC GGAGGCAAAG 238851 GAATAATGAT CATTAAGACA ACGCCCATCA TCCAAAGAGC AAGGATTAAG

238901 TCGCTGGACT TATTGATCAT GTTTAAGGCG GTATCGCCAC CAAGTGTTCT 238951 GCTGACGAAA TTGAGTAGCT TATTCATTAT AAATGATCAG GTTGGTTAGT 239001 ATTTTTATTA TTAGGATTTT GCGCATTCAG TGAAGTGATA TAGAGTAGAA 239051 TTTCTCCAAT AGCTTCGTAA GTAGATTCTG GAATAAATTT TAATTCCTTC 239101 CCTTCATCCA AAAGCTGATG TGCTAAAGGT ACGTTTCGCA TAATGGGAAT 239151 TCCGTACTTT TCAGCTTCAT CAAGTATCCT TTTAGCTCGT AAGTTGATGC 239201 CCATGGCAAT GATCCAAGGT GCTTTATATT TTTCAGGCAT GTAGCCAATA 239251 GCAACAGCAA TATCTTTGGG ATTAGAGACT ACGGTGCTTG CATGTTTCAC 239301 CTGTGATGAC GAGTCTTCAT AGGCAATTTC TTGAGCAATT TGTCGACGAC 239351 GGCCTTTAAT CTCAGGATTT CCTTCCGTGT CTTTAAACTC CTGCTTAACC 239401 TCAAACTTCT CCATCTTTAA TTCTTTAGCG AAATTGTGGC GCTGATAGAC 239451 AAGGTCAAGA ATCGCAACAA TCAAAAAGAA AATTCCTATC GAGGTTACTG 239501 CTTTATAAAA AATTTCTTTG AAGATTTGAG CAGTAATTAT AGGAGAGACT 239551 CCTGCAGTTT CTATAATTAA AGAGACTTTG CTTTTTAACG TTATGTATAA 239601 AATTAAGGCT GCTCCAAAAA TTTTTAAAAT CGATTTGATC AGCTCTATGA 239651 GAGTCTTTAT TTTAAACTTT TGTTTGATGT TCTCAATAGG GTTGAACTTC 239701 TTGATATCTG GTTTAAAAAC TTCGGTAGAA AATGTAGGAC CAACGATAAG 239751 AAAACCTACA ATGACGCCAA CAACAGCAAC AGCTCCCAGT AAGGGAAGTG 239801 ATGCTGTTAA AATAAGCATA AGACAGTTCT TAAGATAAAA TAAGGTAATT 239851 ACAGGATCAT GGCGAGTGGG AGCTTGTGAG AGCATGGAAA CCAGAAAGCC 239901 ACCTAAATGC TTGAAAAAAA AGGTCGATAG GGAGAAAGCĆ GTAAACATAG 239951 AGACGATAAA GGTAACCGCA GAAGGAAAAT CCTGAGATTT TGCTACTTGA 240001 CCTTTTTCC GAGCATCTCT AAGTCGCTTC GGCGTGGCCT TTTCTGTTTT 240051 TTCACCCATG CTGTTGCCAA GGTTCGATTA AGAGGACTAT ATTCGGAATT 240101 ATCTGCAAAT TACCTTATGC TCTTGTTTTT CTCAAGAGTG AAGCCCAAAC 240151 GTTTTGTTCT TAGTAAGATT ATAAATGAAT CTTAGAGATA AAGAAGAATA 240201 GTAGAAAAA AGAACAATAA CGAAAAGTTG TTGAATTTAG CAATACTTTA

240251 GGTCACTTCA AGATGGTTCT ATATGAAGAA AGTATTAGGT AGGGAAAACT 240301 TCAATCGATT ATAGTGGTTT TTTATCATTT TCAAAGCACT CTTGATTTTG 240351 AAGGTACTAT TTTTCTAGCT ATTGTGCTTT TTTGAAACAG AAGAGTTGTC 240401 TTTCTAATAT TCGAAAAAGC ATGTCATTAT TTTTATTTTT AGATGTCTTA 240451 TGAGTCATAC TGAATGTGGA ATTGTAGGGC TTCCTAATGT AGGAAAGTCT 240501 GGCTTATTCA ATGCTTTAAC AGGAGCTCAA GTTGCCTCCT GTAACTATCC 240551 GTTTTGTACT ATCGATCCTA ATGTGGGTAT TGTTCCTGTT ATCGATGAAA 240601 GACTGGAAGC CTTAGCTAAA ATTAGCAATA GTCAGAAGAT CATCTATGCG 240651 GATATGAAAT TTGTAGATAT TGCAGGTTTA GTTAAGGGAG CTTCCGATGG 240701 CGCGGGTCTG GGAAATCGGT TTCTCTCTCA TATTCGAGAA ACTCATGCTA 240751 TTGCTCATGT AGTGCGTTGT TTTGATGATC CAGACGTTAC ACACGTTTCA 240801 GGAAAAGTCA ACCCTGTTGA GGATATTGAA GTTATCAACT TAGAGCTCAT 240851 TTTTTCTGAC TTCTCCTCAG CAAAAAATAT CCATAGCAAA TTAGAAAAGC 240901 TAGCCAAAGG AAAGCGTGAA GTAGGAGCTC TCTTGCCTCT ATTTGATACA 240951 ATTATTGCTC ACTTAGAAAA GGGGCTGCCG CTACGTACTT TAGAATTAAC 241001 TCCAGAACAA ATTGTGGCAT TAAAGCCCTA TCCGTTTTTG ACCATGAAGC 241051 CTATGTTTTA TATAGCTAAT GTTGACGAGA GTTCTCTACC AGATATGGAT 241101 AATGATTATG TTGCCGCTGT TCGGGAAGTT GCTGCAAAAG AAAATTCTAA 241151 AGTGGTTCCT ATCTGTGTTC GTATAGAAGA AGAAATCGTT TCCTTACCTA 241201 TTGAAGAGCG CTTAGAATTT CTTATGAGCT TAGGTCTTGA AAAATCAGGA 241251 CTTCATAGAT TAGTGCGTGC TGCGTATGAC ACTTTAGGAC TGATTTCTTA 241301 TTTTACTACA GGTCCTCAAG AATCTCGTGC ATGGACAGTG GTTCGAGGGT 241351 CTTCTGCTTG GGAAGCTGCT GGAGAAATCC ATACGGATAT TCAAAAGGGC 241401 TTTATTCGTG CTGAAGTGAT TACTTTTGAA GATATGATAG AGTGTCAAGG 241451 TCGTGCAGCT GCTCGAGAAT TAGGGAAATT ACATATAGAA GGACGTGATT 241501 ATATCGTCCA GGACGGTGAT ACTATGCTGT TCCTTCATAA TTAAAGGAAC 241551 CCTTTGCAAA CCAATCTTGA GCATCCAAAA TGTCTTTTTC AATTGCTCGT

241601 ATTAGAGTTT CTTTTGATTG AAACTTTTTT TCTTCTCTAA GAAATTTTCT 241651 CGGGATAATG CTCACTTCTT TGCCGTATAG ATTTTCCGCA AAGGAAAAGA 241701 TATGCGCCTC TGCATATAAA GACTCTCTTC CAAAAGTAGG GGCAGTTCCT 241751 AAATTCATAA CACCCTGACA GGTAGTGCTA TCATAACGTA TTTCACAAGC 241801 ATAAACTCCT AGGGGAATTA AACTTTCTTC TCTAGGAAGA TTTATAGTGG 241851 CGAATCCTAG AGAACCTCCT ATTCCGGAGC CCTCGGTTAT TTTTCCAGAA 241901 ATGGCATAGG GATGACCCAA AAAACGATGA GCACATTCAA GATTCCCTGC 241951 GGACAGAAAC TGGCGGATTG CTTTGCTGGA GACAACTATG TTATCCATAC 242001 GGTAAGGAGG AATCTTGATG ACCTCTATAC CTAACGGCTT GCCTATAGTA 242051 TCGAGAGCCT CGGTATTGCT TTGCTGTTCT TTCCCTATGC AAGAATCATA 242101 ACCTAAGATG AGGCGTTTGC ATTTCAAGTT ACGATGTAAC AAAGTAAGAA 242151 ATTCTTCTGC CGATTGATTC GCAAAATTTA AATCAAAAGT AAGGACACCT 242201 AACCAGTCTA TGGGAAACGT TTGCAATAAT TGGAGGCGCT CTTCTTTTGT 242251 ATTGATGAGT TTCGTGTGAT TTAAAGAAAG TACCGTTTGA GGATGAGAAT 242301 CAAAGGTAAT AACTCCACTG GATCCAGAAT AGGAAGTAAG AATAGATAAA 242351 AGATTGCTAT GCCCTAGATG ACATCCGTCG AAAAAACCTA CAGTTACAGA 242401 ATCTACAGAA AACGAAGACG TTAAACTATA GGCTATTTCC ATGGGCATCT 242451 CGTAGGTAGG GAGAAATATC GAAATCGGGG TGGTCTAATA GATTCCCATC 242501 AATACATTCA TCTATAGAAA AACGGCCACT GCGTAAACGG CGTAGCTGCT 242551 CAAGATAAGC TCCACAGCCT AACATCGTGC CAAGCTCATG AGCAATGCTG 242601 CGAATATAAG TTCCTTTGCT ACAAGAGACT ACAAAATGCA ATAAAGGGTA 242651 CTCATATTTC GTAATCTGCA AGTGAACTTG AACTGTAGAA TGGTGACGTT 242701 CTATAGATAA ACCTTTTCTA GCATATTCAT ACAGCTTTTT CCCTTGGACT 242751 TTTTTAGCGG AAAACATGGG AGGCAGTTGC TGGATCTCTC CTTGGAAATA 242801 CTCGGCAGCT GATAATACTT CTTCGAGACT AGGAATCTTC TTAGATCTTC 242851 CTACAACTTT GCCGTCGCAA TCATAAGAAT CGGTAGTTGT CCCTAAATGG 242901 GCAATTGCTT CGTATTCCTT GTCTTCAAAA AGTAAAATAT CAGAAAGTCT

242951 AGTAAATTTA CGGCCAATCA ACATGACCAT AACGCCAGTA GCGAAGGGAT 243001 CTAAAGTTCC TGCATGACCA ATCTTTTTAA CGCCTATTAA CTTGGTTAGA 243051 GCGCGGATAA GGCTAAACGA AGTTCTCCCT TGAGGCTTGT CTACAAGAAG 243101 AATGCCCTCT TTTAATTCTA CTGCAAGATC CATAGTCATG TCTTTAATAG 243151 TATTCAAGTT CCCAAAAAAT ATAGTTTATT AACTCTTTTC TTTCTCTTGA 243201 ATCTGCCAAA GCAGGTTTTC TATATAATCT TGAGGTGAGA AAATATCATC 243251 GAGATAAAA TGAAGTTCTG GGAAATATTT AAGGACGACA TTTTTCGAAG 243301 CTCTATGAGC GATAAAACCA GCAGAGACTT TTAAAGCTTC TAAAGCCTCT 243351 TCCTTGGTAT TCTCATGAGG CATTACAGAT ACATAAACAC GTGCAGAGTG 243401 CAAATCCTTA GATAGAGAAA CACGAGTTAC CGTGATCCAA AGATTAGAAA 243451 TCTTGGGATG CTTAACATCT TTTAAAATTA CCTTTGCAAT GGCTTCTTGT 243501 AATAAAGCAT TTACCCGTTT AATACGTCTA TTTTCTGTCA TACAGTACTT 243551 CAAGTTATAG TTTTTGTGGA TGATAGATAA CTTCATAACA TTGTAGGACA 243601 TCACCTATTT GAGCTTGCTG GTATCCTTCT AACAAAATTC CACACTCTAA 243651 ACCTTTGCGA ACTTCTTTGA CATCTTCTTT AACACGTTTT AATGAAGATA 243701 ACGTACCTTT CCAAAGGATC TCTTTATTAC GTAATACTCG GACTTTATGA 243751 TTGCGAGTCA TAATTCCTTC AGTAACTATG CAACCGTAAA TAGATCCTAC 243801 TTGTGAAGAC CTAAAGATTT CTTTAATCTC AGCAGAACCT TCATCTTTTT 243851 CTTCAGCAAT AGGATCTAAT AGAGAAGTCA TAATTTCTTT AATTGCATCA 243901 ATAGCATGAT AGATGACGGT AAATAGTTCA ACTCGGACTC CTAAGCTCTT 243951 AATTAAAGGT TCCGCATGAC TTTCTATTCC TGTATGGAAA CCGATGAGAA 244001 CTGCTTTAGA GGCGGCAGCT AAACGAATGT CTGATTCTGA AATTTCTCCT 244051 ACACTGTTTG TTAAAATTTC AACATCTACT TTTTCTGATT TAATCTTAGA 244101 TATTGAACTG ACCAAAGCTT CTATGGAACC TTGAACATCA GCTTTAATCA 244151 TAAGCTTAAG AGTCTTTTTA TTCTGTAACA TAGAATCAAA GTTAGGCCGC 244201 TTCTTTTGCT GTAAAGCAAA ACGCTGTTGT CCTGCGGATC TAGCTTCAAT 244251 AATGTCTCTA GCCGTTTTCT CGTTTTTCAC GACGAAGAAA GGATCGCCAG

244301 CTTTAGGAAT GTCCGATAGA CCTGTGATCA ACACAGGAAT AGATGGCCCA 244351 GCTTCTTTCA TCAATTCATT ATGTTCGTTA TGCATAGTTT TCACTTTGCC 244401 ATAACAATCA TTGAAGACGA GAGCTTCGCC CAGTTTTAAG CTTCCATTTT 244451 GAATCAAAAC AGTCGCAACA GGTCCGAGAC CCTTGTGCAG TTCTGATTCA 244501 ATAACAAGTC CTCGAGCACG TGCTGAAGGA TCGGCTTTTA GCTCCAAGAC 244551 TTCAGCTTGT AAAGCTAACA TCTCTAAAAG TTCTGAAAGA CCTTCTCCTG 244601 TTTTTGCGGA GGTATTTACT GTAACAGTCG AGCCTCCCCA AGCTTCTGGC 244651 AATAGATTGA TTTCAGAAAG TTGTCTATAG ATGGTTTCGG AATTAAAATT 244701 AGGCTTATCA CACTTGTTGA TAGCTACAAC AATAGCGATA TCAGCAGCTT 244751 TTGCATGTTC AATAGCCTCT AAAGTTTGTT CTTTAATTCC TTCGTCTCCA 244801 GCGACTACAA GCACAACAT ATCACAAACT TCAGCTCCAC GGGCTCGCAT 244851 TGCAGAGAAA GCTTCGTGAC CAGGAGTATC TAAAATTGTT ATGTCTCCCA 244901 CTGGGGTGGA GCAGCAGAAG GCTCCCATGT GTTGGGTAAT CGCTCCAGCT 244951 TCTGTTGCAG CGACATTACT TTTCCTTAAG GAGTCAATGA GTGTTGTTTT 245001 TCCGTGGTCG ACGTGACCCA TAAACGCAAC AATAGGGGAG CGAATCACAA 245051 GCTTGCTGGG ATCTGTAGAT TGAATTTCGT CTCTTACAGT GTCATTGCTT 245101 AGGCACAACT TATCTTGCTC AGAATAGTCG ATGTCAATTG TACATCCAAA 245151 CTCTAAGCCA ATAAATTGTA CTGCAGTTTC GCTGTCTAGA ATATCATTGA 245201 CTACATAGGT CATTCCATGA ATGAATAACT TTTGAATGAC TTCTGAAGCC 245251 TTGAGCTTCA TTTCTGCTGC CAGATCTTTG ACGGTAATTG GCAAGGAAAT 245301 TTTGATATGC GTAGGTCGCT GGATAGAGGC TTCGTCATAG TGTTTTTTAG 245351 GCTTATAAAC ACGTTTTTTT CGCCATCTGT CTTCTTCTCC GCCTTCATTT 245401 AATCCGTAAC GATCTCTTCC TGTAAAAGCC TTTAGGCTTT CATCAGATTT 245451 CTTAGAACGA TCACGAAAGT CGGTAAGATT TTTCTTCCCA GCATCTCTTT 245501 TTGGACCAGA AGCAGGACTA CGGTTAGCAG GATTGAATTG TTTTTCTCGA 245551 TTATTCTGTT CACCACCTTC AGATGTTCCA GGTTTCCCTG TTTTATCTGA 245601 AGCAACGGC TTTGTGCTTT TCGAGCCAGC TACGACTTTC TCTTCCTTGG

245651 CAGGAGCCTT GAATGTTTTT GCTAGGAGAT GATTGATATG CTTTCCTGTA 245701 GGGCCGAACT TAGATTTAAT CATTACGACG CTTTTAGGTT CAGCAGGCTT 245751 CACAGGCTTA GGCTCTAATT CTTTTTCTTG AGGTGGGGTT TCGGGCAATA 245801 CGGGTTGCTC AGGAAGAACT TCAGCAACTG GATGAACCTC AGGACTTTCG 245851 TCACAAACCT CATCGACTAC TTCTAACTCA GGCTCAGGAT CTGCTATGGA 245901 GACTGGAGCA GGTTCAGATG TATCCACTGG AATATGAGCA GAAGACTCTT 245951 CTTCGGATGA TGAGAACGAC GAACGATTTT TAGCACGAAT GCGACGTGAA 246001 GTAGACTCTG GTGAAGCTTG TTCCGCACTT GCCGTAGGGG TAGAAGTTGC 246051 GGCAAGAGCT ACTTTTACAG ACTTTTCTTT CGCAGAAGGT TTTTCTGAAG 246101 AAGATTTAGC TTCAGAAGAT CCTGCTTGGG CAAGTTTTTG TTTTAGCTTA 246151 TCCAGCCCCG CGGCTTTCGT TAATTGAGCA TTCTTAATCT TCAATTTCAG 246201 GTTTTCGTC AACTTTACTT TCTCCATATT TGCTGACTTG CTCAAGGATC 246251 TTATAAGCAA GCTCTAAACT GATCCCAGGA ACAGATGCCA GATCATTAGC 246301 ACTCGCTAAT AATACTCTTC TAATTGTGTC ATATCCAGCA TGTTCTAAAT 246351 TTTGGATGAC TAGCTTACTA ATCCCTTCCA TTTCTAAGGG TTGATCTAAG 246401 TGCGGACTAT CGAATTCTGC TAATTGAAGG CGTTGAATTT CTAGCAACTT 246451 ATTGTACTCA CTCATACGTT GTACTTCGAG CTCGTAGTCT AGAATGTGGC 246501 TAATTAAACG AGCGTTAATT CCTCGTTTAC CAATAACAGT AGCGTAGTCT 246551 GCATCATTAA CGACAATTGC AATCACTTTG TCGTCTTCTA AAATAGCAAT 246601 CTTTTGGATT TCTATTGGAT AAAGAAGATT CTGTAATAAC TCTGTAGAGA 246651 CGGGGGAGTA ATTGACAATG TCAATTTTCT CATCGTTCAA TTCTCGAATG 246701 ATATTTTTA CTCGAGAACC TCGCATACCT ACAAAAGCTC CAACAGGATC 246751 AGTTTTAGGG TCTGACGATC TTACAGCTAG TTTCGTGCGG TACCCAGCTT 246801 CACGAGCTAT CTTAACAATC TCCACAGAAC CTTCTTCTAG TTCTGGGACT 246851 TCTTGAATAA ATAATTGTTT AACAAATTCT GCGTGACTAC GACTGAGGAT 246901 AACTTCCGCT CCACCATTTT CAGACTCTTG AACTTCATAG AGTAGGGCGT 246951 AAATTTATC ACCGATCTTA TGTTTTCTG TTTTAGGATA AAACCGGGTA

247001 GGAAGAATTG CTTCAACTTT TCCTAAGTCA ATAATTAAAT TAGAACCTTT 247051 AGCAAAACGT TTGACAACAC CAGATAAAGT TTCATTTACG CGATGGCGAT 247101 ATTCTTCATA AATAACGTCT CTCTCAGCAT GTCTTAGCTT TTGACCGATA 247151 ATTTGTCGTG CTGCGTGAGC AGCTATCTTC CAAAATTATC AGAAACAAAA 247201 GGGACATCCA TGTACTGACC AATCTGACAG TCCGGATCGT ATTCTCTGGC 247251 TTTATCTAAA GGAATTTCTT TGCTAGGATT CTGACAAATT TCTACTATTT 247301 CCTTTTCACA AAAGACTTCG ATGTCACCAG TACGAGAATT AATGTTTACA 247351 GATATGTTCG CGTCATCTCT TAAGGTTTTT TTAGCAGCAA TTTTTAAAGC 247401 AGATTCGATA GCTCCTATAA TAGTAGAGCG CTGAATCCCT TTTTCTTTCT 247451 CCATGTAGTC AAAAATAGCT ACAAGATTTT TATTCATTAT ACTCCTCTTG 247501 TAAATAAGGA AATCAGTAAA TAGCTAAATT AAGGACGTAT TATTTAGAAC 247551 TAAATAATAC GATCCTCTGC ATTACCAGCA TTAGATGCTA TTTTCCTTTT 247601 TTCTTTCTCT CTTTAGGGCC TTGAGAATCC TTGAAATCTA ATTCAGTCCT 247651 AGAGTCTTGA TCATAAGCAT TGTCAGCTAA GTATTCTTTT ACAGAAAGAG 247701 AAACTTTTTT ATGATCTGGA TCTAGCTTAA TTACTTTTGC AGAAACATTT 247751 TCTCCAATGG AGATAATATC TTCAATTTTT GCAAAGGGCT TGTCAGAAAG 247801 TTCTGAAACG TGAATCAATC CTTCAATCCC GTTTTGTAGC TCAACAAAGG 247851 CTCCAAATGC AGTGATTTTA GTCACAACTC CTGAAATTAC TGTGCCAGCA 247901 GGGAACATAG CTTCAATTTC ATTCCAAGGA TTAGAACTTA ATTGCTTAAC 247951 TCCTAAAGTA ATTTTTTAC TTTCTTTGTC TACTGATAAA ATAACAGCCT 248001 CTACAGAATT TCCTTTTTTG AATAGTTCTG AAGGGTGAGA GACTTTTTTA 248051 ATCCAACTCA TGTCAGAAAT ATGAATCAGA CCCTCAATTC CTGGTTCTAA 248101 TTCAACGAAA GCACCGTAAT TGGTTAAGTT CTTGATTTCA GCATTGACAT 248151 GGAGACCTAT AGGATATTTT TCTTCGATAT TGTCCCAAGG ATTACGTTCT 248201 GTTTGCTTTA ATCCTAGAGA AATTTTTCCT TCGTCCTTCT GAATAGATAG 248251 AACAATGGCT TCAACTTCAT CGCCTTTATT TACGACTTCA CTAGGATCTA 248301 CAATATTTT CACCCAAGAC ATTTCAGAAA TGTGAATTAG ACCTTCAATG

248351 CCCTCTTCAA TTTCAATGAA AGCTCCGTAG GGGAGAAGCT TCACAATTTT 248401 ACCAAGAACT CGTTTTCCAG GAGGGTATTT CTTCTCAATA TCTTCCCAAG 248451 GATTATGCTC TTTTTGTTTG AGACCTAGAG CAACTCGTCC TTTTTCTTTA 248501 TCTACGCTTA AAATAATTAC TTCCAACTCT TGATTCAATT CGACCATTTC 248551 GGAAGGATGT CGTATGCGCT TCCAGGTCAT ATCGGTAATG TGGAGAAGAC 248601 CGTCAATACC ATCGAGATCT AAGAATACAC CAAAGTCAGT AATGTTTTTA 248651 ACAACTCCTT TGCGGTATTC TCCGATAGAA ATTTGTTCAA TAAGTTCGGC 248701 TTTCTTAGAG ATTCTCTCAG CTTCTAAGAG TTCTCTTCTT GAGACAACAA 248751 TATTGCGACG TTCAACGTTA ATTTTTAAAA TTTTGAATTC ACAAACTTTT 248801 CCGACATAAT CATCTAAATT TTTGATTTTC TTGTTGTCAA TTTGTGATCC 248851 AGGTAGGAAG GCTTCCATTC CAATATCTAC AATAAGGCCG CCTTTGACTT 248901 TACGTGTAAT TTGACCTTTA ACAATAGAAC CTTCTTCACA ATGAGCTAAG 248951 ATGTATTCCC ATTGACGTTG TCGTGTGGCT TTTTCTCTAG AAAGGACAAC 249001 TTTGCCCTCT TCGTCTTCGG CTTGGTCGAG ATAGACTTCT ACTTCAGCTC 249051 CAAGCACTAA ACCTTCTGAA GAGTCTATGA ACTCTGACAT AGGGATCACT 249101 CCCTCAGACT TCAGACCAAC ATCAACTACG ACAAAGTCTT TATTAATATC 249151 AACTACGGTA CCTTTTAGGA TGGCGCCAGG CTGTATTTCG TTATCAGATT 249201 CTTCTTCGCT CGAAGTAATT CTGTGTGCCG TATAAAGCAA ATCTTTAAAT 249251 TCGGCAACGT CTTCTGTGAG GCATTCTATA TTGTCCAGAA TTTTTTTAGA 249301 TCCCCAAGTA TATTCAGCTT GTTTTGGCAT TTATTGTTAT TCTCCTAAAA 249351 ACTACAAAGT CAAAAAAGAC AGTGTAAATA TAAACCTTAA AAAAGGCAAG 249401 ACCTCCCTTG CAGGAGACGA TTAGTTCTAT ATAATCGAAT TAATCGTAAA 249451 ATCCTATACT TCCGTTCTAG AGATGTATAA GGCCTCTTTT TTAACTTTTC 249501 TATATTACCA GATTTTATAA AATACGTGTG GTAATTATCA TAAATGTGCC 249551 TTGTCATATG TAGAGAAACA GGATTTGGCG CATAGAAAAC ATATCGTTTT 249601 TTTCATAAGC ACCCCATAAA TTCTCGTCCC AGATGGGCTA CATCAGCTTT 249651 GTTCTTAAGC CTGATAGATC TTTTTGGTTG CTGCTCAAGA ACCAAGTGCA

249701 TTCAGAAGTA GAACCATACT TATTGAGGGT TAAATAATAG GATGGGTCTT 249751 GGTTTCTTAG CTCTAAGAAA ATTCTAAAGA TTAATCACAG AAATGTAGGT 249801 GTTATTACTG GAAGATGAAG ATCTTGAGAT TTGAAACTTT AGTCTTAAAT 249851 CCGTATAAAT TAAAACTAGA CAAAAGCTAT TCAAAAGCAA GTTTTTTGCA 249901 TAGAGAAACT GACGCATAGA CGCTAAGTGT TGTTTATTAG ATAGGCTAGA 249951 ATTAAGTGTC TTTTTTCTAG TAGATCAAGA ACAGTGTTCT CAATTTCTAA 250001 AGTTAGATAA GGAAATTATA AATGATTCAT TCCCGGTTAA TTATTATTGG 250051 TTCAGGTCCA TCTGGATATA CAGCGGCAAT TTATGCATCA AGAGCGCTTT 250101 TGCATCCTCT TTTATTTGAG GGGTTTTTCT CTGGGATCTC TGGTGGCCAG 250151 CTTATGACTA CAACAGAAGT TGAGAATTTT CCAGGGTTTC CTGAAGGGAT 250201 TCTTGGGCCA AAACTTATGA ATAATATGAA GGAGCAGGCT GTGCGGTTTG 250251 GGACCAAGAC ACTAGCTCAA GATATTATTT CCGTAGATTT TTCTGTTCGC 250301 CCTTTTATTT TGAAATCAAA AGAAGAAACC TATTCTTGTG ATGCCTGTAT 250351 CATAGCTACA GGAGCTTCTG CTAAACGTTT AGAAATTCCT GGAGCAGGAA 250401 ACGATGAATT TTGGCAAAAA GGAGTGACTG CTTGTGCCGT TTGCGATGGG 250501 TTCTGCTTTA GAAGAAGCTC TTTACCTGAC TCGTTATGGA AGCCACGTAT 250551 ATGTAGTTCA TCGTAGAGAT AAACTGCGGG CTTCTAAAGC TATGGAAGCT 250601 CGGGCGCAAA ACAATGAAAA AATTACATTT TTATGGAATA GCGAGATTGT 250651 AAAAATTTCT GGAGATAGCA TTGTCCGTTC CGTAGATATT AAGAATGTTC 250701 AGACTCAAGA AATTACAACT AGAGAAGCTG CGGGGGTGTT CTTTGCTATA 250751 GGCCATAAGC CAAATACGGA TTTTCTCGGA GGACAGCTGA CGTTAGATGA 250801 GTCGGGCTAT ATTGTGACTG AGAAAGGAAC GTCCAAGACT TCTGTCCCTG 250851 GAGTATTTGC TGCTGGAGAT GTTCAGGATA AGTACTATCG TCAGGCGGTT 250901 ACTTCTGCAG GAAGTGGTTG TATAGCAGCA CTAGATGCTG AAAGATTCTT 250951 AGGCTAATGC GATTGCAGTT GCTGTGGCAT ACTCTTTGCA GTGGCTTATA 251001 GAGAGAATGA CTTTAGAAAT TCCAATTTTT GCATAGACAT GCGAAGGGAG 251051 GAGAACTTCG GGTCCGTGAG ATACTTTAAA GACTTCGATG TCTTTCCAGG 251101 CAACAACGCT CCCTATGCCA GTTCCTAAAG CTTTTGCTAC AGCTTCTTTT 251151 CCAGCAAAGC GACCTGCAAA TGAAGGGATG GGATCGGTCT TTTCTAAGCA 251201 ATATTTCTGT TCTGCTTCTG TAAAGATTCT ATTGAGTAGT CGATTGCCGT 251251 GAGTTGCAAT TGCCTCGCGA ATGCGGCTAA TTTCAATAAT ATCGGTTCCT 251301 ATATGAATGA TTTCCATAGA GTTCGCACTA ATTTCCTTCA GGATTTTCCA 251351 TTAGGGCATT CTGTAATTTT GAGAAGTAAG CGATGGCAGT TTTAATGAGA 251401 GCATAGCCCA CTGTAGCTCG AATCACGAAT AAGGCTTGAG GATTGCTGAT 251451 ATAACGACGA AGGACTAAAG AGAAGGCAAT CATAATGACC ATGATGAGAA 251501 ACCTTCTAGA GATAAAGGTC TGCTTTATAT AGGCAATCCA TGTAATTTTT 251551 TGTGAAAAGA ATTCAGAAGA TAGACTCAGC TGTTTCCTTA TAGTTTTGCT 251601 TAGCAAGTAG CGATGTTTTA GGGATGCTAA TCCCCAAGAA ACCATAGATA 251651 GAAGACAGTA GGTAAAAAAA GAAGAAAGGG GGTCTAGGGC TATTGCGGTA 251701 GCTTTTAGCA AGAGTTTCAT TCCTGCTGAT ATCCACAAAA TACCAGGAAA 251751 TAGTATCAGG AAATATTTGA TGTTTCTTGC CATAGTGCAT TACACTTAAT 251801 TTCTAAATAC TCTATTTTAC TTGGAGTATG AATTTTTTGC TAACTGGTAT 251851 TAAGATATGC GAGCTGAGAT GGCTGTGATC TATTGGGATC GCTCAAAAAT 251901 TGTCTGGTCT TTCGAGCCAT GGTCTCTAAG ACTTACTTGG TATGGCGTCT 251951 TTTTTACTGT AGGGATTTTT CTAGCATGTC TCTCAGCAAG GTATTTGGCT 252001 CTTTCCTATT ATGGTTTGAA AGATCATTTA AGTTTTTCCA AAAGCCAGCT 252051 ACGCGTGGCT TTAGAAAACT TTTTTATATA CTCTATTTTA TTTATTGTCC 252101 CTGGAGCTAG ACTTGCCTAT GTGATTTTTT ATGGATGGAG TTTTTACTTA 252151 CAACATCCTG AAGAGATCAT TCAAATATGG CACGGAGGCT TGTCGAGTCA 252201 TGGAGGCGTT CTTGGCTTTC TTTTGTGGGC GGCCATTTTT TCTTGGATAT 252251 ATAAAAAAA GATTTCAAAA TTGACTTTTC TCTTCCTTAC AGACTTGTGT 252301 GGATCAGTTT TCGGAATTGC AGCGTTTTTT ATTCGTTTGG GTAATTTTTG 252351 GAATCAAGAA ATTGTAGGAA CACCGACTTC TTTGCCTTGG GGGGTGGTTT 252401 TTTCTGATCC TATGCAAGGT GTCCAGGGAG TTCCTGTGCA TCCTGTGCAG 252451 CTTTATGAGG GAATCAGTTA CTTGGTCGTC TCTGGAATTT TATATTTTCT 252501 TTCCTATAAG CGTTATTTGC ATTTAGGTAA GGGATATGTG ACTTCTATAG 252551 CCTGTATTTC TGTCGCTTTC ATTCGTTTTT TTGCGGAGTA TGTAAAAAGC 252601 CATCAAGGGA AAGTTTTAGC AGAGGATTGT CTACTTACAA TTGGTCAAAT 252651 TTTATCTATA CCTTTATTC TATTTGGTGT GGCCTTACTT ATCATTTGCT 252701 CATTGAAAGC TCGAAGGCAC CGTTCACACA TATAGCAGCA TCTTTTTGGC 252751 ATGAGACTTT AGAATTTTAC AGTCTATCAC AATAAGAGAC CTGTATTGAG 252801 AGATTCTCCT TTGGGAGATA ATGTGGGTTT ATGATATTTT GACCTCTTAA 252851 GTTTAATTTT TAGAGTTTAT CAAATGAATA AACGCACTTT GCTTTTTGTT 252901 TCTTTAATTG GGATTGCTTT TGTAGGATGT CAAATATTTT TTGGTTATAA 252951 TGAATTTCGT TCCTGCAAAA ATCTAGCAGA GAAACAAAGA AAGATTTCAG 253001 AACAGACGCT AGCTGCAGTA GAATCTGTAG GGTTAAGTGT AGCTTCATGG 253051 GACACCGATG TAAACGGAGA AGAACATAAG AATAACTACG CAGTTCGTGT 253101 TGGAGACAAG TTATTTTAT TACATAATGG AGAAGCTGCT CAGTCTGTTT 253151 ATTCTTCTGG GGAATCTTGG AGCTTTGTAG ATCACAAGTG TGGTTTCGAT 253201 AATATTCACT TGGCTTTATA TCGTCAGCAG GGTTCTTCTT TCAATCCTAC 253251 GAATACGGGA AAAGTTTTTC TTCCTACGAA TCATGAAGGT TTACCTGTAC 253301 TAGTTGTTGA GTTTCGTAAC AATAAAGAGC CTTTAGTATT TCTAGGTGAG 253351 TACGCACAAG GAAGAATTTC CAATAAAGAT AGCACGATCT TTGGTACAGC 253401 GCTTGTCTTT TGGAGATCAG GAAGCGACTA TATTCCTTTA GGTCTCTATG 253451 ATTCTCGAGA AGAAAGTTA GTTTCTTTGG ATCTTCCTAT TACACGAGCT 253501 GTAATTTTG GTAATGACCA AGATTCGGCA AAGTCGTCAG ATACTGCGAA 253551 CCACTATGTT TTATTTAATG ATTACATGCA GATTATTGTT TCTGAAGAGA 253601 GTGGTTCTAT AGAAGGTATC AATTTACCTT TTGCTTCAAC AAATAATAAA 253651 AGCATTGTGA ATGAAATTGG TTTTGATAGG GATTTAGCTT CAGAGAAATC 253701 TCCTGAAGCT CTTTTCCCTG GGCTGTCTTC AAAACTTCCT GATGGCCAAC

253751 AAGCCAAAAA CTCGATTGGA GGTTACTACC CTTTATTGCG CAGGGGATTA 253801 TTAAGTGATT CTAAGAAATT ACTTCCTCTA GAGTATCACG CATTAAATGT 253851 GGTTTCAGGA AGAGAGCTAG CGACTCCTGT GGCTTTAAGA TACCGAGTTC 253901 TTTCCTATAC CCCCCATTCC ATTCAATTGG AAAGCTTAGA TAGATCGGTT 253951 CAGAAGGTAT ACAAACTTCC AGAGAATCCG GAAGAAAAGC CCTATGTTTT 254001 TGAAACTGCA ATTACTTTAA CGAAAGAAAC CGAAGATGTA TGGGTAACTT 254051 CAGGAGTTCC TGAAGTGGAG ATCATGTCAA ATGCTTCAGC CCCAACCATT 254101 AAATACAGGG TTATCAAAAA AAATAAGGGG TCTTTAGATA AAGTTAAGCT 254151 TCCAAAAGTA AAAGAGCCTT TAGCTATACG TCGTGGTGTT TATCCTCAAT 254201 GGATTTTAAA TTCGAATGGA TATTTCGGTA TTATTTTAAC TCCGTTGTCT 254251 GAAATTGCTT CTGGCTATGG ATCTCTCTAC ATTTCGGGTT CTACGGCTCC 254301 GACAAGATTG TCTGCTATTT CTCCTAAAAA TCAACTGTAT CCAGTATCAA 254351 AATATCCTGG ATATGAGACC TTGCTTCCTT TGCCAAAAGA TGCAGGGACA 254401 CATCGATTTT TAGTGTATGC AGGTCCCTTG GCAGAGCCTA CACTTAAAGT 254451 ATTAGATAAG ACAATTACTC AGGAGAAGGG AGAAAATCCT GAGTATCTTG 254501 ATAGCATTTC TTTCCGTGGT GTTTTTGCAT TTATTACAGC TCCTTTTGCA 254551 GCACTCCTAT TTATTATTAT GAAGTTCTTC AAATTGGTTA CGGGTTCTTG 254601 GGGAATTTCC ATTATTTTAC TTACTGTATT TTTGAAATTG CTTCTCTATC 254651 CTTTAAATGC ATGGTCCATA CGATCTATGA GGCGTATGCA GATTTTATCT 254701 CCTTATATTC AGCAAATTCA GCAAAAGTAT AAGAACGAAC CTAAGCGTGC 254751 TCAGATGGAA ATCATGGGCT TGTATAAGAC AAACAAGTG AATCCTATCA 254801 CGGGTTGTTT ACCTTTATTG ATACAGCTTC CTTTCCTAAT TGCGATGTTT 254851 GATTTATTAA AGTCATCATT CTTATTACGA GGAGCCTCGT TTATTCCTGG 254901 GTGGATTGAT AACTTAACAG CTCCTGATGT GTTGTTTTCT TGGCAGACAT 254951 CGATATGGTT TATTGGAAAT GAGTTCCACT TACTTCCTAT TCTATTAGGT 255001 ATAGTGATGT TCTTACAACA GAAGGTCACG AGTTTGCATA AGAAAGGACC 255051 TGTTACGGAT CAGCAGAAAC AGCAACAAGT TATGGGGAAC ATGATGGCGA

255101 TTTTATTTAC CGCTATGTTC TATAACTTCC CTTCAGGATT AAACATCTAT 255151 TGGCTTTCGT CTATGATTTT AGGAGTCGTC CAGCAGTGGA TCACTAATAA 255201 GATATTAGAT AGCAAACATC TTAAAAATGA AGTGGTTTTA AATAATAAAA 255251 AACATCGATA ATACTGAAAA AAAAGTCGAA TTGTCTTAAT ACAGGCTTAT 255301 TGAAACAAGC GATTTATATG AGCTCATAGT TTATGCGAGC ATGGGAAGAA 255351 TTTCTTTTGC TACAAGAGAA AGAAATTGGC ACAAATACTG TAGACAAGTG 255401 GTTGCGATCT TTAAAGGTCT TATGTTTTGA TGCTTGTAAT TTGTATCTTG 255451 AAGCTCAAGA TTCTTTCAA ATTACTTGGT TTGAGGAGCA TATAAGACAT 255501 AAGGTTAAAT CTGGTCTTGT AAATAATAAC AATAAGCCCA TTCGTGTTCA 255551 CGTTACTTCG GTAGATAAAG CAGCTCCTTT TTATAAGGAG AAGCAGATGC 255601 AGCAAGAGA GACAGCATAC TTTACCATGC ATTATGGAAG TGTGAATCCT 255651 GAGATGACCT TCTCTAATTT TTTAGTTACC CCTGAAAATG ATCTTCCTTT 255701 TCGTGTTTTA CAGGAATTTA CTAAGAGTCC TGATGAAAAC GGAGGAGTTA 255751 CTTTTAATCC AATTTATCTG TTTGGACCTG AGGGATCTGG AAAAACTCAC 255801 TTAATGCAGT CAGCTATCAG TGTTCTTCGT GAATCTGGAG GTAAGATTCT 255851 CTATGTTTCT TCGGATTTGT TTACAGAGCA CTTAGTCTCT GCTATCCGTT 255901 CAGGAGAAAT GCAAAAATTC CGTTCTTTTT ACCGCAATAT TGATGCTCTA 255951 TTCATTGAGG ATATCGAGGT TTTTTCAGGA AAGTCGGCAA CTCAAGAAGA 256001 GTTCTTCCAT ACGTTTAATT CTCTTCATTC TGAAGGGAAG TTGATTGTAG 256051 TGTCTTCATC CTATGCGCCT GTGGATCTCG TTGCTGTTGA AGATAGATTG 256101 ATCAGCAGGT TTGAATGGGG AGTTGCAATT CCGATACATC CTTTGGTTCA 256151 GGAAGGATTG CGCAGTTTCT TAATGAGACA GGTAGAGCGC TTATCTATTC 256201 GCATTCAAGA AACGGCCTTA GATTTTTTAA TTTATGCGCT ATCTTCCAAC 256251 GTAAAGACCT TACTGCATGC ACTGAATCTT TTAGCAAAGA GGGTAATGTA 256301 TAAAAAACTC TCTCACCAAT TACTATATGA AGATGATGTG AAAACTCTTT 256351 TAAAAGATGT TTTAGAAGCA GCAGGAAGCG TTCGTTTAAC TCCTTTAAAG 256401 ATTATTCGTA ATGTTGCTCA ATATTATGGG GTCTCTCAGG AGAGTATTTT

256451 AGGACGTTCT CAGTCCCGAG AATATGTATT GCCACGTCAG GTAGCCATGT 256501 ACTTTTGTCG TCAGAAGCTT TCACTATCAT ACGTGAGAAT AGGCGATGTC 256551 TTTTCAAGAG ATCATTCGAC GGTAATCTCA TCCATACGAT TGATTGAACA 256601 AAAAATAGAA GAAAATAGCC ATGATATTCA CATGGCTATT CAAGATATTT 256651 CTAAGAATTT AAATTCCTTG CATAAGAGTT TGGAATTTTT CCCTAGCGAA 256701 GAGATGATTA TTTAGAGGAG TAACGATCCT CTAATTTTTG CTTCTGGACA 256751 GTTGCAGCAG CCGGCGCAGA TATCATTTTC ATTGAGTTCA ACTCCGCTGT 256801 TCTAGCTTGT TTAAATTGTC TAATATGATT GAAAGCAAGG ACAAGAGCAA 256851 TACCAGCAAG TACTATCTGT ATAGCAAGAA CAATGCCGAT TCCTAAGGAT 256901 AGGGTGACTC CACTGGCAAT AACCAGTCCT AGGGTGGTTA AGGAAGCAAG 256951 GAGGAGACCT AGACTAACTA AAATAGGCTT AAAAAATACC CAGGAACAAT 257001 AACGATGTGT TGCCTTGTGA GATACTGATG GTTGTGGTTG TGTAGTCTGA 257051 GGTGTTTGTG CTACTGTAGC CATGTTGTTC TCCTGAGTAA AAAATTAAAT 257101 AATTTTAATC TGTCTATATA ATTGTTGTCA ATAATATTTT TATAGATATT 257201 TAGTTCAGAA CTATTAAGCT CTCTTAATAG GCTCTCTTT ATTTACCAAG 257251 GCGAAGTTCT TTTTTTTTT TCAAATTTCG CATATGGAGG GTTCTAGATG 257301 GTATTTGAAA TGGTTGCATT GTGGAAGATT TTTCGAGTTT TGATAAGAAC 257351 AAAGTCAGTG TTGACTCTAT GAAACGGGCG ATTTTAGATC GTCTGTATTT 257401 AAGTGTTGTA CAATCACCAG AGTCCGCATC TCCTAGAGAT ATCTTCACAG 257451 CTGTTGCAAA AACTGTTATG GAATGGTTGG CCAAGGGGTG GCTGAAAACT 257501 CAAAATGGCT ACTATAAAAA TGATGTAAAA AGAGTTTATT ACCTTTCCAT 257551 GGAATTTCTC TTAGGGAGAA GTCTAAAAAG CAATCTTTTG AATTTAGGAA 257601 TTCTAGATTT AGTAAGGAAG GCACTAAAAA CTTTAAATTA TGACTTTGAC 257651 CACCTTGTAG AAATGGAATC CGATGCAGGA TTAGGAAATG GTGGTTTGGG 257701 GAGACTGGCA GCTTGTTACT TGGATTCTAT GGCTACATTA GCAGTTCCAG 257751 CCTACGGCTA CGGTATACGC TATGATTATG GTATTTTTGA TCAGAGGATC

257801 GTCAACGGGT ATCAAGAGGA AGCTCCTGAC GAGTGGCTAC GTTATGGAAA 257851 TCCTTGGGAA ATCTGTAGGG GAGAGTACCT CTATCCCGTA CGATTTTATG 257901 GAAGGGTCAT TCATTATACC GATTCTCGAG GGAAACAGGT GGCAGATCTT 257951 GTCGATACCC AAGAGGTATT GGCGATGGCT TATGATATTC CGATTCCTGG 258001 GTACGGTAAT GATACTGTAA ATTCTCTAAG GCTATGGCAA GCACAATCTC 258051 CGCGAGGCTT TGAATTCAGC TATTTTAACC ACGGGAACTA TATCCAGGCT 258101 ATAGAAGATA TCGCCTTGAT AGAAAACATC TCTCGCGTCC TCTATCCTAA 258151 TGATTCTATT ACTGAGGGGC AGGAATTGCG TCTCAAACAA GAGTATTTTT 258201 TAGTTTCAGC AACCATTCAA GATATTATCC GCAGATATAC AAAGACACAT 258251 ATTTGTTTGG ATAACCTTGC GGATAAAGTC GTAGTACAAT TAAACGATAC 258301 CCATCCCGCT TTAGGGATTG CTGAAATGAT GCATATTTTA GTCGATAGGG 258351 AAGAATTACC TTGGGATAAG GCTTGGGAGA TGACTACAGT CATCTTTAAC 258401 TATACCAATC ATACAATCCT CCCAGAGGCT TTAGAGAGAT GGCCTCTCGA 258451 TTTATTCTCT AAGTTATTAC CTCGGCATTT AGAGATTATT TATGAAATAA 258501 ATTCCCGTTG GTTAGAAAAA GTTGGCTCTC GCTATCCTAA AAATGATGAT 258551 AAGCGCCGGT CTTTATCCAT TGTTGAAGAA GGGTATCAAA AGCGTATCAA 258601 TATGGCAAAC CTTGCCGTAG TAGGTTCTGC AAAAGTAAAT GGAGTTTCGT 258651 CATTCCACTC TCAGCTGATT AAAGATACTC TCTTTAAAGA GTTTTATGAG 258701 TTTTTCCCTG AGAAGTTTAT CAATGTGACC AATGGGGTGA CTCCACGACG 258751 ATGGATTGCT CTCTGTAATC CTCGTTTGAG TAAGCTTCTC AATGAAACTA 258801 TAGGGGATCG TTATATCATT GATCTTTCTC ATCTTTCATT GATCCGTTCC 258851 TTTGCCGAAG ATAGTGGTTT CCGAGATCAT TGGAAAGGGG TAAAATTAAA 258901 AAATAAGCAG GATCTAACAA GTAGAATTTA TAATGAAGTT GGAGAAATAG 258951 TAGACCCTAA TTCTCTCTTT GACTGTCATA TTAAGCGTAT TCATGAGTAT 259001 AAACGACAAC TAATGAATAT TTCTTAGAGT CATCTATGTT TATAATGACT 259051 TGAAAGAAAA CCCTAATCAA GATGTCGTCC CTACAACAGT AATTTTTTCT 259101 GGTAAGGCGG CTCCTGGCTA TGTCATGGCC AAACTCATTA TCAAGTTAAT

259151 CAATAGCGTT GCTGACGTTG TAAATCAAGA TTCTCGAGTT AATGATAAGC 259201 TTAAGGTTCT TTTTTTACCT AACTATCGAG TTTCTATGGC TGAGCATATC 259251 ATTCCTGGTA CAGATCTTTC AGAACAGATT TCTACAGCTG GAATGGAGGC 259301 TTCTGGAACA GGAAATATGA AATTTGCTTT GAATGGAGCT CTGACTATAG 259351 GAACTATGGA CGGTGCAAAT ATAGAAATGG CAGAGCATAT TGGTAAGGAG 259401 AATATGTTTA TTTTTGGTCT TTTGGAGGAG CAAATTGTAC AACTGCGGAG 259451 GGAATACTGT CCTCAGACAA TTTGTGATAA GAATCCTAAG ATCCGTCAGG 259501 TTTTAGATTT GCTAGAACAG GGATTTTTCA ATAGCAATGA TAAAGATCTG 259551 TTTAAACCGA TAGTACATCG CCTACTGCAT GAAGGAGATC CCTTTTTTGT 259601 CTTGGCTGAC TTGGAGTCTT ATATCGCTGC CCATGAAAAT GTGAACAAAC 259651 TCTTTAAGGA ACCAGATTCA TGGACTAAGA TTTCTATTTA TAATACTGCA 259701 GGAATGGCT TTTTCTCTAG TGACAGAGCC ATTCAGGATT ATGCCAGAGA 259751 TATTTGGCAT GTTCCTACAA AATCTTGCTC TGGAGAAGGA AATTAAGAAA 259801 TAGACAGGAT GGAATCAGGG ACTCTTTCCA TAGCCAAGGA GCTATAGAAA 259851 GAGTCCTTTT TGTTCAAAGA TTGCTAGTTT AATAGTAGGA CAGCCGGAGC 259901 TTCTAAGATC TTTTGTAATC GTTTCATAAA CATCGCAGCA GGATAACCAT 259951 CAATCACTCT ATGATCTACA GATAGGGTAA GATTGCAGGT AGATCCTATA 260001 GTAATTTCTC CGTCAAGAAC AAGAGCTTGT TCTGTAACAC TTCCTACGGC 260051 AAGAATCGCC GCTTGAGGAG GATTGACAAT CGCTGTAAAT TCAGTGATTC 260101 CTGTCATTCC TAAGTTAGAG ACACAGAAGG ACCCTCCTTT GTATTCAGTG 260151 TCTTGAAGAG ATTGATTTCT TGCTTTTAAC GCTAAGCTCT TAATTTCTGC 260201 TGAAATCATG CCGAGATTTT TACGGTCTGC GCAGCGTATA ATTGGCGTAA 260251 TAATTCCATC TGGAATGGCC ACAGCTATCG AGATATCGAT AGTATCAAAA 260301 CGGACGATTT TATTATCGAC ACTGTTAAAT CCTGAATTGA TAGAAGGGAA 260351 CTCTTTGAGC GCCAGAGCAC AGGCACGTAC AATGCAATCG TTAATAGAGA 260401 GTTTGATTCC CTGAGCTTGA AGTTCTTTGA GCAGATTAAG GAGAGGTGAG 260451 GCGTAGACCT GCTGCCTTAC ATAGAAGTGA GGAATAGAGA TCTTAGCAGC

260501 TTGTAGGCGT GCAGCAATCA CTTCCCGAAT CGGAGAGAGA TTCTCCTCAT 260551 GATAGGAACC TGGAGGCACT TCGGGAGACT CAGGATAGCC AAAACCAGCA 260601 ATGCTTTTAG GAGGAGCTTT CTCTAAATCT TTTTTTACTA TACGTCCTCC 260651 AGGACCACTC CCTTGAATTG ATGAGACATC TATGTTTTTC TCTTTTGCTA 260701 GTTGTCTAGC TAATGGAGAG AGATTATTCG TAGTGCCTAC GTGTTTGAAG 260751 ACTAAAGGCG AGGAGAGAG TGGCTCTGGC TTAAAAGTTA CTGCTGTGAA 260801 TGTTGCTGAG GCAGCTTGTG GAGTTGTTGC AGGCGAGACC TCTTCAGAAG 260851 AACCTTTTGG AGATGCTTCA AGGTTAGAAG GTTCTGTCTT AGGAAGAAGT 260901 TCTTCTAGAT TAAAGGGCTC GTTGGCTTCT GTAGAGAGTA CCGCAATAGG 260951 GGTGCCTATA ACGATTTTCT CGCCTTCATG ACGTAAGATT TCACGAATCC 261001 AGCCATCTTC ATTTGCTGTA TGTTCTAAAA TAGCTTTGTC TGTAGAGATC 261051 TCTACAATGA CGTCTCCAAA ACTGACCTGA TCATTACTTT TTTTATGCCA 261101 TTTCACTATA GTGCCCACTT CCATAGTTGG AGAAAGCTTT GGCATTTTCA 261151 ATAAGGAGAT CACAAACTTA CCTCATGACT TTTTCAATGG TATCTAAGAT 261201 TCGGTTAACA TTAGGCAAAG TGGCCTGTTC TAAGATTTTA CTATAGGGCA 261251 TAGGCGTTTC TTTTTGGCAT ACCCTTAAGG GGGGAGCATC AAGAGAATCA 261301 AAAACATGCT CAGTAATCAG GGCAATAATT TCAGAAGAAA TCCCAGCGAA 261351 GTAGTGGCCC TCTTCAATTA CAATACAGCG TGAAGTTTTT CGTACCGATG 261401 ATAAAATTGT TGATATGTCT AAAGGTTTGA TCGTTCTTAG ATCAATAATT 261451 TCTATAGACA AGCCCCAACG TTTTTTGGCT AGAGAACACG CTTCTTTTGT 261501 AATGGAAACC ATACGGCTAT AAGTAATAAT TGTAAGGTCA TTTCCTTCTT 261551 GAACTCTATG TGCTTTCCCA ATAGGAACGA GATATTCTTC GGTGGGGACT 261601 TCCCCTTTTA AGTTATATTC TAGCTCGTTT TCTAAAAAAA GAACGGGGTT 261651 ATTATTTCTG ATTGCTGATT TTAATAAGCC TTTAGCGTCG TAAGGGTTCG 261701 AAGGGGCTAT AATAATAAGA CCTGGAATAT TAGCATACAA CGACTCAACG 261751 CAATGAGAAT GCTGGCAAGA TACCTGGGCT GCAGCACCAT TAGGGCCACG 261801 AAAAACTATA GGAACGGAAA ACTTCCCTCC AGTCATAAAA TGCATCTTAG 261851 CTGCATGAGA AATGATTTGG TCCAAGGCTA CAAAGGAAAA GTTCCAGCTC 261901 ATAAATTCTA TAATAGGGCG CAGGCCTGAC AATGCGGCTC CTATTCCAAT 261951 TCCAGAGAG GCTGCTTCAC TAATAGGAGC ATCAATGACT CTCTTAGGGC 262001 CCCATTTATC TAATAAGCCT TTGGTGACTT TATAAGCACC ATTGTAGTCA 262051 CCAACCTCTT CACCAAGAAT ACAGACATTA GGATCGCGAG ACATCTCTTC 262101 GTCAATTGCT TCTCGGAGAG CTTCTCGAAT TTCTAATGTT TTATGTTTAG 262151 GCATAGACTC CTTCCTCTAA TGTGGTGACG GATGGATCTG ATGAGAGTTT 262201 TGCGTTAGAG AACGCTTCTA AAACAGCAGT TTTGCATTCT TGGCGTATAT 262251 TTTGAAATTC CTCTTCAGTC AGAACCTCTA ATCGAATTAG CCAATCTTTA 262301 GCTAGGACAA TAGGATCTTT TTTAAATAAA CACTGCATTT CTTCTTTCGA 262351 TCTATATAAA TTAGGATCTG ATATAGAATG CCCTCGAAAT CGGGAGCAGA 262401 GACACTCAAC TAAAACCGGA GATTCGGTAT CAACCATATA GCGATAAGCC 262451 TCTCTAAATC CTAAAAGAGA GTTAAATAGA TCAAAACCAT TGACTGTGAC 262501 TGCACGGATA TCGTAGGAAC TTCCTTGAGA CTCTGCTATG GGCTGTTTTG 262551 CAACAGCACG ATTTAATGAC GTTCCCATAC TCCAGCCGTT ATTTTCAATA 262601 ATAAGCATTA GAGGGAGTTG GTGAAGAGA ACAAAGTTCA GAGTTTCATG 262651 GAATACACCT TGAGCTACCG CACCATCTCC GATAAAGCAT AGAGAAACTC 262701 TATTTTTTG TTCTTGATAT TTGATGGTAA ATGCGGCTCC AGCTGCGAGG 262751 GGAATTTGTC CTCCGACAAT ACCAAATCCT CCAGGGAAAT TAGGCCCACA 262801 CATATGCATG GATCCTCCAC GACCTAAAGC GCATCCAGTT TCTTTCCCTA 262851 AAAGTTCAGC AGCAATTTCT TGAAGGGGAA TGTTGAGAAG AATCGCAAGT 262901 GCGTGGCAGC GGTATGAAGA GAACACCCAG GGATCTAGTC CTGTGTTTGC 262951 GATTGCAGCA GTTGCTACAG CTTCTTGGCC AGCGTAAGAG TGGTAAAATC 263001 CACCCACTAG CCCTTCTAGA TAGGCTTCTT CTCCTCGGGC TTCGAATTCA 263051 CGAATCAGAA CCATCTGTTT TAAAAATTTA ATACAGGAAG CGGGCCCGTA 263101 AAGGTCTAAG ATCCTTTCTA CTGTGGATTT CTCTGTGCCC TGAGAAGCTA 263151 TATTATAAGG TGCTGAACTA TCCATAACTT TTTTATAAGG GAGACGCTTC

263201 GGGAGCGGTT TTTGATCTCA TTATAGGAGC CTAGGTTTCA AAAGACAAGG 263251 GCTATCCGAA GAAAAAATAA TTAAGTTTTT TTAAAAAAATC GAAAAAATTT 263301 TAATATGTTT CTCGTGGCAA TTTAAAGAGA ACAAAAAGGA TTACATCACA 263351 TGGACGTTTC TCGTAAAATT AATCGACACA CTCAGTTTTA CGTGGATTCG 263401 ATAGACGTG TCATCAAAAA CTTTGATCAT AAGCCTAGTG AAGATAAATC 263451 TCGAGATCAT GAAGAATTAG AGGAAAAGCT TTTAACTATA ACAAAACGTA 263501 TTGTTGCTTC TGCTCAAGAA TTTCAGAATC GTAAGACGGA CTCTAAAAAC 263551 TACTACTTAA AAAAAACTCA ATGGTTGCCA TTCAAAAATG AAGAGTTGGA 263601 GCAAACTAAG GAATTGTTTG CCATGTTAAC TTCGATGGAT AAAAAAATAG 263651 CTCAGTTATT TTTTTATTCT CCCGGATGTA GTAGCGACTG GGTAGAATTT 263701 ACCGAAGTAA TTTGTCATTT AAATGATTCT ATAGGTTTGG GAGGGGTTTT 263751 ACTGTGTTGC GGATTATTCG AACAACAGTG TGAGCATGTT GTTACTGTGA 263801 ATAAAAAGCT AGACCTACCG CTTCTTTTAG GAACTACCGT TGTAAATAGT 263851 CTACGTTATT ATTTAACCTA TAGGAATATC TCTCTTTTGA ATTGTCAGAG 263901 CATGAGTGAG CTAGGCAAGG AGCTCGGGGA TGTTTTAAAG CAACATGGAG 263951 TTGCTTTCAC TTTAATTTTT AAAGAGATTG TGGATATAGA TTTATTAAAC 264001 TATGTGAAAC TAATTCAAGG ACTCAAGCGT AGTGGAAATA TTCAAGCACG 264051 TATTTATGAT AATGATGTGC CTACTTTACC TTCAGTTTCA TCGAGTCCTA 264101 TTGCGTTGAG ATATAGCTTA GCGAATACAA TTCGGGGCCT TGCTTTACAT 264151 GTAGATTTTT CTTCATTGAA GTTTATAAGC CCGTCTATAC TTTCCAATAC 264201 AGAGCACACT GCAAAAGCTT TAAACTCTGG CGGAGAATGT TTTATCTTTT 264251 CTAATCTGGA TGAATTCAAT CTTGGAATGA AAATAGTCAT GCAACTATTG 264301 CGGACAGGGA AGATTAGCCC AGAAATTTTA AATAAAAATA TCATGAAAAT 264351 TCTCATGATA AAAAGAAGAG TACGTTCTTT ATATATTTAA TTATTTGGTC 264401 TTTTTTTGA TTTCAAAACT AATAAAATAA ATAGAATTTT AGATCTCTGA 264451 AGGAATCTCA GCAAGGCTGG GAGTCGATAG ATCTCTTACT TGTTTTTCTA 264501 ACTTACTTAG TCTTTCTTCA GTTTTAGGAA GGTTCCGAAT TTTAGCAATC

264551 AACCGATGTG TTTCTTGATA AGGTCGTGCT GGAGCGCCTC CATAAATGCC 264601 TGGAGAGGTG ATAGATTTTG TGACTCCAGT TTGAGCAATC ATGATCACAT 264651 GGTCTGCAAT AGAAATATGC CCAGTAATTC CGGTTTGCCC TCCAATGATG 264701 ACATGTTCAC CAATTTTTGT AGAACCTGCA ATGCCTGCTT GGGCAACAAT 264751 AATACTATGC TTTCCAATTT CTACGTGATG AGCTACTTGT ACTTGGTTAT 264801 CTATTTTAGT TCCTTCATGG ATCACGGTGT TCTTGAATCG ACCACGATCT 264851 ATCGTAGTGT TGGCTCCGAT TTCTACATCA TCACCTACAA TCACATAGCC 264901 TAGATGCTTT AAAGGTTTGT GATGACCAAA AGCATTTGTA ATATAACCAA 264951 AACCACAGGA TCCTAAAACA GCTCCAGGTT GAACAACTAC ACGGTTTCCC 265001 ATGAGGACTC TTTCTCGAAT CACCACCTTA GGGTGAATCA GACAGTTAGC 265051 ACCTAGAACG CTGTGAGCTC CAATGACACT TCCAGCTCCG ATGTATGTGT 265101 CAGAGCCGAT ATGGGCATGT TGACTAATGA CAACGTAAGG TTCTATGGTT 265151 ACATTTTCT CAATACGTGC AGTAGGATGA ATCACTGCAG TAGGATGAAT 265201 ACCAGGAAAC CCTGATGTTA CGGGTTCAAT AAACAACTCT ATGCACTTTT 265251 GAAATGTTAG AGAAGGGGAT TCATTGGTAA TAAGAAAGTT TTTCTTTAGG 265301 TGGGCATGTT GCATTGCCTG AGATCTAGAT AAAATAATAG CACCAGCTTT 265351 GGTGTTTTTT AGAAAGCTAG AGTATTTCTC ATTATCTAAA AAAGCAATAT 265401 GGTGAGGTTG CGCCTGACTA ATATCTTCAA CACCTGAAAT AGGAGTTTCT 265451 ATATTTCCTT GAACTTCGAC TTGTAGTAGC TCAGCTAACT GTTTAAGAGT 265501 GTAGACTGGT GCTTCGGACA TAGAAAACTC CTTAAACTTG GACTAGTTTT 265551 GTTTTTTGAA AGATTCGTTA AGAATAGCAA TAATTTCGGT TGTTTTATCA 265601 GTCCCAGGTG CTATTGCTAA GACAGCTTCT TCATTAAGGA TAGCTTCTAG 265651 TTTTTCTTTG GACCGCACTG ATTCTGCAGC TATTTTTACT TCTTGAATGA 265701 GTTTTTGAAT GCGTTTTACA TGTACTTTGA TTGATAGATT GATAGTACTG 265751 AGACTGGTAC GCATTGTACT CTCCTGAAAG ATCTTCGAAT TTCTTTCGCA 265801 ACTCTTCAGA GGCAGAATCC GATAGGCTTT CCATGTAATC TTCATCTTGC 265851 AACTTATTAT AAATAGAAGT GAGTTCTTCT TCTATTTTCT CAGCATTTTT

265901 TACAAACTGC TGTTTCATAG CTTCCAATTC TTCAGTTTCC TTTTTACCTA 265951 GATCGGATTC TTCAAGACAT CGCTTTAAAT TAACATAGCC TAAATTTGCA 266001 TGAGCTGCGC TTGTTGATCC TAAAACAAGA AGAAATGTAG AAAATAATAA 266051 TTTTTCATA ATTACCTGTA CTAACTTGGG TATAGAAAAA GGGTACCAAA 266101 AGCCTTTTCT GAAAACAACA AAGATTTCCT TCGATAAGTC CTTAATTTAT 266151 ATCTTAGAAC ATGCCCCCTA AAGCAAAGAA GAATCGCTGA GATACATCAA 266201 TTTTTCTCC ATTCAAAGTC TCGGTTGGAC GGAAGGGCCA ACCAAATCCT 266251 AACATAACAG GAACATTATT CATTACATCG AAGCGCAGAC CAAATCCAGC 266301 ACTACTACGT AGATCTTTTA ACGAAATCTT ATACTCTTGT AAACCGACAA 266351 AACCTGAGTC TAAGAATACA AAGGCACTAA TATTAGGTTG TCTGATGAGA 266401 GGGTATTGAA ACTCTTCTGA AATAAGGAGC GAAGAGAGTC CTCCCTGAGG 266451 TTCTGTAGCA GAGTATTTTG GACCGATAAT AAAGGATTTA TATCCCCGAA 266501 CTGTAGTCTC TCCACCTAGG AAGAAGCGCT CACTGACAGG AACTCCTTCA 266551 GCTGTAGTAT TGCTATAGGG TTTAATAAAT TGAGCTTCCC CTTTGATTTT 266601 CAAAATACCT TTACGCGTAA GTTTTCTATA GATAGAGCTG TTTAAAGAGA 266651 GTTTTGTAAA ATGATAAGTT CCTCCCAAAC CAGAAACCTC AAAAGTCACC 266701 CCCCGCGAA TCCCTGTAGT TGGAGTTCTA GGACTATCTA CAGAATCGTA 266751 ATTCAAGTTG ACACCTGCAG CAGAGACAAA TCCTTTATTG CTGTCTATAT 266801 TTGGCCCTAG GAGGAACTTA CGTTTTTCAT GTAAACTCGT TTGACTTCCT 266851 CGATAAAATA GACCGTATTT CAGGTGTTCG TTCAAGATAT ACGTTGTGCT 266901 GACGTTCCCG CCATAGGTTT GGACAGCATA ATCTTTAGAT AATGCTCTGT 266951 TAATTGATTT ATCTAATTCA ATTCCTAAAA TCCAAGGAGT GTTTAGAAAA 267001 TGAGGTTTGG TCCACTTCAA AGTATAGTCT GTGACTTTGT CCCCGAAGTT 267051 GGCTTTTAAG AATAGATGTT CTCCACCGCC TCTTAGACAA CGAAAACCTT 267101 TAGAAAATAT ATTTCTAGCT CCAAATAGAT CAAAATTACT TTCAGATAGT 267151 TCAATTCCTC CAAAAAGATT GTCAAGAGAA CTAAATCCTA AGAATAAGCC 267201 TAAGTTTCCT GTTGTTGTTT CTTTGACTTC TACAAAAATA TCTCGGTATT

267251 GATCCGCATT GCCCATAGGA TCAAGTTGAG AACGAACTGT ATAGACACTA 267301 ACGCTTTGGA AGTAGCCTGT ATTTCTTAAA CGTTGCTCAG TATCTTCTAG 267351 CTTTAAGCGA TTGAATGTAT CTCCTGGGAA GAGACTGGTT TCGTGTAAAA 267401 TAACGTCAGA TTTTGTATGG GTATTCCCAG TAATTTTAAT TAACCCAACT 267451 TTATAAGGAG ACCCTTCACT TACCTCATAA GTTACATCAT AAATAGGGCG 267501 GGTTGCGTGA GGGATGAAGA GAACGTCTAC ATTGGTATTG ATGTAGCCAT 267551 ACTTTGCATA AGTTTGTTTG ATCTTATGAG CCCCATCCCA TATTTTATCG 267601 GGGCAATAAA GATCATTGGG GCCGACTTGG GATTGCTTTT CTATAAGGCG 267651 TTTTGGCAAA ACCTCAAACC CTTGGATATG GACGTGTCCT AAGGTATATC 267701 GCGACCCTCG ATCAATATCC ATGTAAAGAA GAATATTCCC TTTGTCGTCA 267751 AGGTCATAGT GAGAGTTGAC TATAGCATCA GCGTACCCGT TATTATGTAG 267801 GTAATTCGTA ATTGCCAAGC TATCTTGTTC AACAATATCT GGGTGATAGA 267851 GTCCAGCTCC AGTAAACCAA CTTGTAGTTG TAGAGTGCTG CTTGGTTTGA 267901 ATAAATTCTT GGATATCTGA TTTTTCTGAT CGAGAGATTC CTGAGAACGT 267951 AAGCTGTTTA ATTTTCCCGC AAGGACCTTC ATTGATTTTA ATTAAAACAT 268001 CGATGTGACC TTTTTCTTGA TTGTGTTCCA GACTGTAGTC TACACTGGAT 268051 GCGAAATATC CTCGCTTGAG ATAATACGTT CTTAGATCAT CAAGACCCTT 268101 AAGAAATTTT TCTCGTTCAA AGAGATCATT ACGGTAAATT TGTAGGGTTT 268151 TAAGAATTTT ATGTTCAGGA ACGACTTGAT TTCCTGAGAT ATGAATATTT 268201 CGAATTGAGG GTTTAGCTAT TAGGTGAAGG GCTATGTTAG TTTTCCCTTC 268251 AGAAAATTCT ACTTTAGGCT CAACAGAGTC GTATTCTTTA GCTAGAATTC 268301 TCAAGTCTTC ATCAAAATCT AATTGAGAAA AAAGAGCCCC ACTTCTGGTC 268351 TTTAATTTGG GTAAGGGATG TTTATTTGAA GCATTTTCTC CTTCCGTTAT 268401 GATTGTGATA GAGTCTACCA CCACATGGCC TTCTTTAACT TTTTCAGTAG 268451 AAAATAAAGT TAAAGGGGTT TGGATTAACG CTAGAATAGA TATTTGCAAG 268501 ATAACTTTAT TTCGCATGAT GAGCATTCCC AAGAGTCTTC CCTAGATAGA 268551 AAGCTTGTTT ATCAGATAAT GAGGTAAGCA CAAAAATAGG GACAAAGAAG

268601 TTTTTTAACA AGACCTTACT ACTTAGATTA TTTTTGTATT TCTAAAGAAA 268651 TTTAAATTGT GTTTGATTTT TTCCCGTCCT AACACCGGGA TTATTACTGT 268701 GCTCAGACTA GAGCCTGAAT ATTTCGGCGG GTTTTTTGCT TCCTTAAAAA 268751 GCTTAATAAA CCTTGTCGCT CAAATAACTC TACTGATTAC ATGGTTTTAA 268801 GGCAAATTCG ATAATTCACT GTGTCCTAGG GGAAAGTAAG AAAGACTAGC 268851 ATAGAAGAGA ACCTTGTTTA CTTTTAGGAT AAGAGAGCTG CTAATAGGAG 268901 TGTCGTCCAG AAAAAGCTCT TGCCAGTGTC CCTGAATCTA CATAATCAAA 268951 AGATAAGCCT ATAGGAAGAC CTAAAGCTAG ACGGGAAATA TTTACAGAGA 269001 AATGTTGTAA TTCTTGTTTT AGAAAAAGGG CAGTAGCATC TCCTTCTAAG 269051 GTTGCATCAA TGGCTAGGAT AATTTCTTTT GGGCATAGCG TTTCTATGCG 269101 TGATTTTAAA ATGGAGAGAC GCTCGTTTTC TATATGTTTC CCTGTAATGG 269151 GCGATAAGAG TGAACCAAGA ACATGATAAC GTCCCTTGAA TACTTTAGAA 269201 CGTTCTAGAA AGAAAACATC TTTTGGAGAA GCGACAATAC ATAGACTTTG 269251 GTTATCTCTT TCTTCTCTAC AAAAGTGACA GTCTGCCTCT TTAGATTCTT 269301 TGAGAGTAAA ACATAGGGGA CAGTGACTAC GCTCACTAGC AACATTATGA 269351 AAAGCGTTAC CTAATATTTT TAATTGTTCG CTGTCCCAAG AGATGAGTTC 269401 AAAAGCAAGT TTTTCTGCTG TTTTAAATCC AATTCCTGGA AGTTTTCGTA 269451 AAAAGAAAAT TAATTTAGAT AAGTAATCTG GATATCTTGT CATAGTGATG 269501 TGAATTTAT TTTTACATTC GGGTCTGGGG CCTAAATTAA GGTTAGAGTA 269551 TAAACTCTCT GAATAGTATA CTAGCTTTTT TCTTTACATG TGGTTCTCTG 269601 TGAATAAAAA CAAAAAAGCA GCAATTTGGG CAACGGGTTC CTATTTGCCT 269651 GAGAAAGTTC TTTCAAACGC AGATTTAGAA AAAATGGTAG ATACCTCTGA 269701 TGAGTGGATC GTGACCAGAA CGGGGATCAA AGAGCGTCGT ATTGCTGGAC 269751 CTCAGGAGTA CACTTCTCTT ATGGGAGCCA TCGCTGCAGA GAAAGCTATA 269801 GCAAATGCGG GTTTAAGCAA GGATCAGATT GACTGTATCA TTTTCTCGAC 269851 AGCAGCACCA GATTATATTT TCCCATCAAG CGGAGCTCTT GCTCAAGCAC 269901 ATTTAGGCAT TGAGGATGTC CCTACATTTG ATTGCCAGGC GGCTTGTACT

269951 GGGTATTTGT ATGGTTTGTC TGTAGCTAAG GCTTATGTAG AATCAGGTAC 270001 ATATAACCAT GTATTGTTAA TTGCTGCTGA TAAGTTGTCT TCTTTTGTAG 270051 ATTATACAGA TCGGAATACC TGTGTGTTGT TTGGAGATGG AGGAGCTGCT 270101 TGTGTCATAG GGGAGAGTCG GCCAGGATCT TTAGAGATTA ATAGGTTGTC 270151 TTTAGGCGCA GATGGTAAGC TAGGAGAGTT ATTAAGCCTT CCTGCTGGAG 270201 GTAGTCGTTG TCCTGCTTCT AAAGAGACTT TACAATCAGG CAAACATTTT 270251 ATTGCTATGG AGGGAAAAGA AGTTTTTAAG CATGCTGTGA GACGTATGGA 270301 AACGGCAGCT AAACATTCGA TAGCCCTGGC AGGCATTCAG GAAGAGGATA 270351 TAGATTGGTT TGTACCTCAT CAAGCTAATG AAAGAATAAT AGATGCTTTA 270401 GCGAAGCGTT TTGAGATTGA TGAGTCTAGA GTGTTTAAGA GTGTACATAA 270451 GTATGGAAAT ACTGCGGCCT CGTCTGTGGG CATTGCTTTG GATGAATTAG 270501 TTCATACAGA ATCCATTAAG CTTGATGATT ATTTACTTTT AGTTGCCTTT 270551 GGGGGCGGTT TGTCTTGGGG CGCAGTAGTT TTAAAGCAGG TCTAATAAGG 270601 ACGATAATTT CATGAAAAAA CGTTATGCTT TTTTGTTCCC AGGACAAGGG 270651 AGCCAATATG TAGGTATGGG ACAAGACCTA TATATGGAGT ATCCTGAGGT 270701 TAGAGAGCTT TTTGATTTTG CTAATGAAAG GTTAGGATTT TCTCTGACTT 270751 CAATTATGTT TGAAGGTCCT GAGGATCTTT TGATGGAAAC AGTACATAGT 270801 CAGCTAGCTA TTTATCTTCA TAGCATGGCT GTGGTAAAGG TTCTATCTCA 270851 GCGTTCTTCT ATTCAGCCTT CTTTAGTCTC TGGATTAAGT TTAGGGGAGT 270901 ATACTGCTTT AGTTGCTTCC GATAGAATCT CCGTGCTCGA CGGCCTTGAG 270951 CTTGTTAGAA AGCGTGGTCA GTTAATGAAT GAAGCTTGTA ATCAGAGCCC 271001 AGGGGCTATG GCGGCTTTAT TAGGGCTTCC CTCTGAAGTT ATAGAGGAAA 271051 ATATAACAAG TCTTGGTCAA GGAATTTGGA TTGCTAATTA TAATGCACCC 271101 AAACAGCTTG TAGTGGCTGG AATAGCAGAA AAAGTAGACC AAGCGATTGA 271151 GTTATTTCGT GATTTAGGAT GTAAAAAAGC AGTTCGTTTA AAGGTGTCTG 271201 GAGCATTTCA TACTCCTTTA ATGCAAGTTG CTCAAGATGG CTTAGCTCCA 271251 GACATTTATG CTTTATGCAT GAAAGATTCT AGCCTTCCCT TAGTGTCACA

271301 CGTGGTAGGA AAATCTTTAG TAAATACTGA AGAAATGCGA GAGTGTTTAG 271351 CTCGGCAAAT GACATCACCT ACGTTATGGT ATCAGAGTTG TTACCATATC 271401 GAATCAGAGG TGGATGAGTT TTTAGAATTA GGTCCAGGAA AAGTTTTGGC 271451 TGGTTTAAAT CGCTCTATAG GGATTTCTAA ACCGATTACA AGTCTTGGTA 271501 CTTTTGCTCA GATTGAAAAA TTCCTATCAG AGGTATGATT TGTATGGATA 271551 TAACATTAGT AGGCAAAAAA GTTATAGTAA CTGGAGGATC TCGAGGAATT 271601 GGACTCGGGA TAGTTAAGCT TTTTCTTGAG AACGGAGCAG ATGTAGAAAT 271651 TTGGGGATTG AATGAGGAGC GAGGTCAGGC TGTTATAGAA AGTTTAACAG 271701 GCTTGGGTGG CGAAGTTTCT TTTGCTCGTG TGGATGTGAG TCATAATGGT 271751 GGAGTGAAAG ATTGCGTGCA GAAATTTTTA GATAAGCACA ACAAAATAGA 271801 TATTTTGGTA AATAATGCAG GCATTACCAG GGATAATTTG TTGATGCGTA 271851 TGTCTGAGGA CGACTGGCAA TCGGTGATTA GCACCAACTT GACTTCCTTG 271901 TATTATACAT GTTCCTCAGT GATTCGCCAT ATGATTAAGG CGCGTTCAGG 271951 ATCTATTATA AATGTGGCTT CTATTGTTGC TAAGATCGGT AGTGCGGGCC 272001 AGACCAACTA TGCTGCTGCT AAAGCTGGGA TTATTGCTTT CACAAAATCT 272051 TTAGCTAAGG AAGTAGCTGC AAGAAATATT CGTGTCAACT GCCTTGCTCC 272101 AGGCTTTATT GAAACAGACA TGACAAGCGT GTTGAATGAC AATTTAAAAG 272151 CTGAGTGGCT TAAGTCGATC CCTTTAGGTA GGGCTGGCAC TCCAGAAGAT 272201 GTTGCTCGTG TGGCGTTGTT TTTAGCCTCG CAGTTATCGA GCTATATGAC 272251 CGCGCAGACA CTGGTTGTTG ATGGGGGGATT GACTTACTAA GACAATAGAA 272301 GAAAGGGATT TGAAAATTTC TCTTCGAGAA CTAATTAAGT AACCGTCGAA 272351 TAAAAAATGA TTTTTTGCGA TACTAATTCT CTTTCTCTTT GTCCCTAGGG 272401 AATAGTGAAG TATTGTATAG TTTAAATAGT AAAAGGATAT AAGCAATGAG 272451 TTTAGAAGAT GATGTAATAG CAATTATTGT TGAGCAGTTA GGAGTGGATC 272501 CAAAAGAAGT TAATGAGAAC TCTTCTTTTA TTGAAGACTT GAATGCTGAT 272551 AGTTTAGATT TAACAGAATT GATTATGACT TTAGAAGAAA AATTTGCTTT 272601 TGAAATTTCA GAAGAAGATG CTGAGAAGCT TCGTACTGTC GGGGATGTAT

272651	TTACTTATAT	TAAGAAACGT	CAAGCTGAAC	AATAAACTTT	СТТАТАТТСТ
272701	GGGTGCGTCG	TAATCTTTTT	TAAAGCTATT	GTTTTTCAAT	AGTGATTACG
272751	GCGTGCCTTT	TTTTCTATAG	AGTAGCTCAT	CTAGAAAGAT	TTATTTTGTC
272801	TTTTTAAGGT	TCTCTGAACT	TGATTTGTTT	AGCATAGAGC	TCTAAAAAAG
272851	ATAAAGCTAC	GGATGGGCAC	TCTTCCACAA	TGTTTAGAAT	TTGTCCTTTG
272901	CTAAGAACTA	GCATGCGGAC	TTGTGTATTT	GCAGAAGCAT	TGTATTCCCT
272951	GGGCTTATTA	TTGAATAAGC	TTTCCTCTCC	AAAACAATCT	AAAGGTTTTA
273001	AATTTAGAGG	AGACTCTAGT	TTTTCTTTAG	AGATCGTAAT	GTATCCTTCT
273051	ACAATGATAT	AAAAGCTGAA	TCCAGGTTGT	CCTATAGAGA	ATACATTGCT
273101	GCCAGGCTTA	AATATTATCG	TTTCAGTTTT	ATCGGCAATT	GTTAAAAGAA
273151	GGTCCATGTC	TAAAGATTGG	AATATAATCG	TTTTTTTTAG	TAGAAAGGCG
273201	CGATCGATCA	AATTCATAAA	AAAGTTCCTT	ATTCACACCA	TAGTTTTAGT
273251	TTTT				